



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

DETEKCE LOGOPEDICKÝCH VAD V ŘEČI

DETECTION OF LOGOPAEDIC DEFECTS IN SPEECH

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. MILAN PEŠEK

VEDOUcí PRÁCE
SUPERVISOR

Ing. HICHAM ATASSI

BRNO 2009



**VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ**

**Fakulta elektrotechniky
a komunikačních technologií**

Ústav telekomunikací

Diplomová práce

magisterský navazující studijní obor
Telekomunikační a informační technika

Student: Bc. Milan Pešek

ID: 89276

Ročník: 2

Akademický rok: 2008/2009

NÁZEV TÉMATU:

Detekce logopedických vad v řeči

POKYNY PRO VYPRACOVÁNÍ:

Navrhněte a realizujte v prostředí Matlab metodu, která bude schopna detekovat chybnou výslovnost v řeči. Prozkoumejte různé metody výpočtu příznaků a klasifikace pro tento účel. Výsledkem práce má být softwarový nástroj pro detekci logopedické vady hlavně u dětí pomocí počítače.

DOPORUČENÁ LITERATURA:

- [1] Psutka J., Muller L., Matoušek J., Radová V., Mluvíme s počítačem česky. Academia Praha 2006.
- [2] Psutka J., Komunikace s počítačem mluvenou řeči. Academia, Praha 1995.
- [3] R. Duda, P. Hart, D. Stork, Pattern Classification, druhé vydání. Wiley, 2003.
- [3] Zaplatílek K., Doňar B., Matlab-začínáme se signály, nakladatelství BEN 2007.

Termín zadání: 9.2.2009

Termín odevzdání: 26.5.2009

Vedoucí práce: Ing. Hicham Atassi

prof. Ing. Kamil Vrba, CSc.

Předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

Abstrakt

Tato práce se zabývá návrhem a realizací softwarového nástroje pro detekci logopedických vad v řeči. Jelikož je třeba odhalit logopedické vady v řeči co nejdříve, je tento nástroj zaměřen na mluvčí dětského věku.

Úvodem text popisuje teorii vytváření řeči, modelování vytváření řeči pro její číslíkové zpracování, fonetiku, logopedii a základní logopedické vady v řeči. Dále jsou popsány použité metody pro extrakci příznaků, pro segmentaci slov na hlásky a pro klasifikaci příznaků do tříd vadné a správné výslovnosti. V závěru textu jsou uvedeny výsledky testování vybraných metod.

K rozpoznání logopedických vad v řeči jsou použity algoritmy pro extrakci příznaků MFCC (Melovské keprální koeficienty) a PLP (Perceptivní lineární predikce). Segmentace slova na hlásky je provedena pomocí metody sledování rozdílů příznaků. Extrahované příznaky hlásky jsou klasifikovány do tříd vadné nebo správné výslovnosti jednou z testovaných metod rozpoznání vzoru. Pro klasifikaci příznaků jsou testovány metody k -NN (Algoritmus k -nejbližších sousedů), SVM (Algoritmy podpůrného učení), ANN (Umělé neuronové sítě) a GMM (Smíšené Gaussovy modely).

Klíčová slova

Řeč, logopedická vada, extrakce příznaků, segmentace slov, rozpoznání vzoru, PLP (Perceptivní lineární predikce), MFCC (Melovské keprální koeficienty), k -NN (metoda k -nejbližších sousedů), SVM (Algoritmy podpůrného učení), ANN (Umělé neuronové sítě), GMM (Smíšené Gaussovy modely).

Abstract

The thesis deals with a design and an implementation of software for a detection of logopaedia defects of speech. Due to the need of early logopaedia defects detecting, this software is aimed at a child's age speaker.

The introductory part describes the theory of speech realization, simulation of speech realization for numerical processing, phonetics, logopaedia and basic logopaedia defects of speech. There are also described used methods for feature extraction, for segmentation of words to speech sounds and for features classification into either correct or incorrect pronunciation class. In the next part of the thesis there are results of testing of selected methods presented.

For logopaedia speech defects recognition algorithms are used in order to extract the features MFCC and PLP. The segmentation of words to speech sounds is performed on the base of Differential Function method. The extracted features of a sound are classified into either a correct or an incorrect pronunciation class with one of tested methods of pattern recognition. To classify the features, the k -NN, SVN, ANN, and GMM methods are tested.

Keywords

Speech, logopaedia defect, feature extraction, segmentation of words, pattern recognition, PLP (Perceptual linear prediction), MFCC (Mel-frequency cepstral coefficients), k -NN (k -nearest neighbors algorithm), SVM (Support vector machines), ANN (Artificial neural network), GMM (Gaussian Mixture Models).

Bibliografická citace práce

PEŠEK, M. *Detekce logopedických vad v řeči* . Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 71 s. Vedoucí diplomové práce Ing. Hicham Atassi.

Prohlášení autora o původnosti práce

Prohlašuji, že svou diplomovou práci na téma „Detekce logopedických vad v řeči“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne

.....

podpis autora

Poděkování

Děkuji vedoucímu diplomové práce Ing. Hichamovi Atassimu za účinnou metodickou, pedagogickou a odbornou pomoc a další cenné rady při zpracování mé diplomové práce. Další poděkování patří PaedDr. Lence Němcové za poskytnutí logopedických nahrávek a firmě APS s.r.o. za podporu celého projektu Děkuji rovněž svým přátelům a hlavně své přítelkyni za pomoc a podporu při tvorbě této práce. V neposlední řadě děkuji rodičům za podporu během celého studia.

Obsah

Seznam obrázků.....	10
Seznam tabulek.....	11
Úvod.....	12
1 Vytváření řeči a modelování vytváření řeči.....	13
1.1 Proces vytváření řeči.....	13
1.1.1 Dechové ústrojí.....	13
1.1.2 Hlasové ústrojí.....	13
1.1.3 Artikulační ústrojí.....	14
1.1.4 Modelování procesu vytváření řeči.....	16
2 Fonetika a logopedie.....	17
2.1.1 Fonetika.....	17
2.1.2 Logopedie.....	18
3 Analýza řečového signálu.....	21
3.1 Lineární prediktivní (předpovědní) analýza.....	21
3.1.1 Perceptivní lineární prediktivní analýza (PLP).....	22
3.2 Homomorfní zpracování řeči.....	24
3.2.1 Melovské keprální koeficienty (MFCC).....	24
4 Segmentace slov na hlásky.....	29
4.1 Metoda sledování rozdílnosti příznaků.....	29
5 Klasifikace příznaků řečového signálu.....	31
5.1 Algoritmus k -nejbližších sousedů (k -NN).....	31
5.2 Algoritmy podpůrného učení (SVM).....	32
5.3 Smíšené Gaussovy modely (GMM).....	33
5.4 Umělé neuronové sítě (ANN).....	34
5.4.1 Model neuronu.....	34
5.4.2 Matematický popis neuronu.....	35
5.4.3 Umělé neuronové sítě.....	35
5.4.4 Typy neuronových sítí.....	35
5.4.5 Algoritmus učení backpropagation.....	36
6 Detektor logopedických vad v řeči.....	39

6.1	Předzpracování.....	39
6.2	Segmentace slova a výběr hlásky	40
6.3	Klasifikace	41
6.3.1	Trénování klasifikátoru a tvorba testovací a trénovací databáze.....	43
6.4	Učíme se správně mluvit.....	44
7	Testování algoritmu pro výběr hlásky ze slova	47
8	Testování klasifikátorů	50
8.1	Výběr dat.....	50
8.2	Extrakce příznaků	50
8.3	Klasifikace	51
8.3.1	Přehled nastavení parametrů jednotlivých klasifikátorů:.....	51
8.4	Výsledky testování klasifikátoru k -NN.....	52
8.5	Výsledky testování klasifikátoru SVM.....	60
8.6	Výsledky testování klasifikátoru ANN.....	60
8.7	Výsledky testování klasifikátoru GMM	62
8.8	Celkové výsledky testovaných klasifikátorů	64
	Závěr	67
	Použitá literatura	68
	Seznam zkratk	69
	Seznam symbolů	69
	Obsah přiloženého CD	71

Seznam obrázků

Obr. 1.1	Model vytváření řeči [2].	17
Obr. 3.2	Postup výpočtu kepra pomocí DFT [2].	24
Obr. 3.3	Blokové schéma MFCC.	27
Obr. 3.4	Princip PLP.	28
Obr. 4.1	Princip segmentace slova na hlásky.	30
Obr. 5.1	Příklad k -NN klasifikace.	31
Obr. 5.2	Příklad SVM klasifikace použitím lineární funkce.	32
Obr. 5.3	Příklad SVM klasifikace použitím funkce RBF.	33
Obr. 5.4	Model neuronu.	34
Obr. 5.5	Schéma neuronové sítě Back-propagation.	36
Obr. 6.1	Princip navrženého detektoru.	39
Obr. 6.2	Příklad rozdělení slova na segmenty.	41
Obr. 6.3	Příklad výběru hranic hlásky	41
Obr. 6.4	MFCC a PLP keprum pro správnou a vadnou hlásku [r].	42
Obr. 6.5	Hlavní nabídka programu	44
Obr. 6.6	Procvičování hlásky [š].	45
Obr. 6.7	Vyhodnocení výslovnosti hlásky	46
Obr. 8.1	Blokové schéma testování klasifikátoru.	50
Obr. 8.2	Grafy vycházející z tabulek 8.1 až 8.4.	58
Obr. 8.3	Graf nevyšší úspěšnosti k -NN pro jednotlivé hlásky, $k = 3$.	59
Obr. 8.4	Graf úspěšnosti k -NN v závislosti na kombinaci příznaků.	59
Obr. 8.5	Graf nevyšší úspěšnosti ANN pro jednotlivé hlásky.	61
Obr. 8.6	Graf úspěšnosti ANN v závislosti na kombinaci příznaků.	62
Obr. 8.7	Graf nevyšší úspěšnosti GMM pro jednotlivé hlásky.	64
Obr. 8.8	Graf úspěšnosti GMM v závislosti na kombinaci příznaků.	64
Obr. 8.9	Celková úspěšnost klasifikátorů k -NN, ANN a GMM.	65
Obr. 8.10	Úspěšnost klasifikátorů pro jednotlivé kombinace příznaků.	65

Seznam tabulek

Tab. 6.1	Přehled počtu nahrávek v databázi	44
Tab. 7.1	Výsledky testování automatické segmentace pro 36 slov.	48
Tab. 8.1	Výsledky testů k -NN pro hlásku [r].....	53
Tab. 8.2	Výsledky testů k -NN pro hlásku [ř].....	54
Tab. 8.3	Výsledky testů k -NN pro hlásku [s].....	55
Tab. 8.4	Výsledky testů k -NN pro hlásku [š].....	56
Tab. 8.5	Výsledky testu k -NN pro $k = 3$	57
Tab. 8.6	Výsledky testu ANN	61
Tab. 8.7	Výsledky testu GMM.....	63

Úvod

Děti s vadami řeči, především s takzvanou dyslalií (patlavostí), přibýlo za poslední desetiletí skoro o padesát procent. V roce 1997 se u logopedů léčilo přes 90 tisíc lidí. V roce 2006 měli v evidenci už téměř 130 tisíc pacientů. Devadesát procent z nich jsou děti a mladiství do devatenácti let. Další tisíce dětí „přeléčují“ vyškolené učitelky v mateřských a základních školách. Tato statistika může být zkreslená tím, že se řečovým vadám věnuje více pozornosti než dřív. „Nárůst artikulačních poruch řeči u jinak zdravých dětí je však jednoznačný,“ říká logopedka Jitka Kaulfussová viz [1].

Cílem práce je vytvořit nástroj pro automatickou detekci logopedických vad v řeči, který by pomohl odhalit logopedické vady v dětské řeči. Práci logopeda však nemůže nikdy zcela nahradit. Měl by mít podobu jednoduchého, pro dítě příjemného, grafického prostředí, kde by se zobrazovaly obrázky a dítě by tyto obrázky slovně popisovalo. Program by vyhodnotil jeho výslovnost, například v podobě smějící se nebo mračící se postavičky. Zároveň by program mohl vytvářet statistiku úspěšnosti dítěte, která by byla dále k dispozici logopedovi.

Automatická detekce logopedických vad je velice rozsáhlá problematika. V této diplomové práci jsou použity pouze některé vybrané metody z velké škály metod pro zpracování řeči. Tyto metody jsou dále testovány z hlediska vhodného použití.

V první kapitole je vysvětlen proces vytváření řeči a modelování tohoto procesu pro číslíkové zpracování řeči.

Následující kapitola se zabývá základními pojmy z fonetiky a logopedie. Důraz je kladen především na logopedickou vadu řeči, tzv. dyslalií, která je nejrozšířenější odchylkou ve vývoji řečových schopností.

Třetí kapitola popisuje analýzu řečového signálu, konkrétně metody PLP a MFCC, které byly dále použity pro extrakci příznaků ze signálu.

Ve čtvrté kapitole je uveden algoritmus pro segmentaci slova na hlásky pomocí metody sledování rozdílnosti příznaků.

Kapitola číslo pět se věnuje metodám rozpoznání vzoru, konkrétně klasifikátorům KNN, SVM, ANN a GMM, které byly testovány pro klasifikaci příznaků řečového signálu.

Následující kapitola popisuje samotný návrh a implementaci algoritmu pro detekci logopedických vad v řeči.

Výsledky testování navržených algoritmů jsou obsahem kapitoly sedmé a osmé. Následně jsou tyto výsledky zhodnoceny a je vybrána metoda s nejlepšími výsledky pro konečnou aplikaci.

1 Vytváření řeči a modelování vytváření řeči

Řeč je přirozenou a nejčastěji využívanou formou komunikace mezi lidmi. Mluvená řeč se přenáší komunikačním kanálem ve formě akustických vln, tzv. akustického signálu. Podstatou akustického (nebo též řečového) signálu je vlnění elastického prostředí v oboru slyšitelných frekvencí. Komunikačním kanálem se rozumí prostředí, kterým je akustický signál, tj. řeč, přenášen od svého zdroje, úst řečníka, až k příjemci informace, uším posluchače. V akustickém signálu řeči je zakódováno několik druhů informací. Vedle samotné akustické složky (amplitudově-frekvenčního spektra) bývá z hlediska komunikace nejdůležitější informace lingvistická, daná například svojí fonetickou, morfologickou, syntaktickou, sémantickou či pragmatickou strukturou, protože vyjadřuje význam sdělované myšlenky. Akustický signál dále obsahuje jisté specifické informace o mluvčím, které respektují charakteristiky hlasového traktu řečníka a způsob artikulace, včetně případných anomálií, jako jsou právě vady řeči [2].

1.1 Proces vytváření řeči

Pro vytváření řeči existuje v lidském těle několik skupin orgánů, které se souhrnně nazývají řečové (artikulační) orgány, či jen mluvidla (artikulátory). Z hlediska tvorby řeči tyto řečové orgány tvoří hlasový trakt. Hlasový trakt lze rozdělit na tři základní ústrojí: dechové, hlasové a artikulační [2].

1.1.1 Dechové ústrojí

Dechové ústrojí představuje hlavní zdroj energie pro řeč. Je umístěno v hrudním koši a tvořeno přírodní dýchací cestou, plicemi a s nimi funkčně spjatými dýchacími svaly (bránicí). Při nádechu dochází k pohybu vzduchu, který tak poskytuje zdroj energie pro řeč. Při výdechu potom v plicích vzniká výdechový proud vzduchu, který je v zásadě základním materiálem pro tvorbu řeči. Výdechový vzduch je z plic odváděn průdušnicí (tracheou), prochází hrtanem a nadhrtanovými dutinami, kde se modifikuje, a jako řečový signál je vyzařován rty do okolního prostoru. Rychlost s jakou vzduch opouští plíce, je přibližně konstantní. Trvání výdechu má vliv na to, jak dlouhý úsek řeči lze vytvořit bez přerušení. Síla výdechového proudu ovlivňuje způsob fungování hlasového ústrojí a tím má vliv na sílu hlasu a částečně i na jeho výšku. Další nádech pak opět doplňuje vzduch do plic, čímž mimo jiné dodává nový materiál pro tvorbu řeči. Navenek se nádech projeví jako pauza v jinak souvislé řeči [2].

1.1.2 Hlasové ústrojí

Pojmem hlasové ústrojí se také často označuje celý systém řeči. Zde se pod tímto pojmem rozumí pouze ta část, kde dochází k samotnému vzniku hlasu. Hlasové ústrojí je uloženo v hrtanu, který je s plicemi spojen průdušnicí. Z hlediska tvorby řeči nejdůležitější část hlasového ústrojí tvoří hlasivky, které se nacházejí v hrtanové dutině přímo za „ohryzkem“. Jsou to dvě ostré slizniční řasy, které vedou napříč hrtanem v místě jeho nejužšího průchodu. Jejich typická délka je asi 15mm pro muže a 13mm

pro ženy. Z jedné strany jsou napojeny na chrupavky hlasivkové a z druhé strany na chrupavku štítnou. Jsou pokryty sliznicí a jejich základ tvoří hlasový vaz a hlasivkový sval. Prostor mezi hlasivkami tvoří hlasivková štěrbina trojúhelníkového tvaru. Jestliže člověk mlčí, pak hlasivky drží hlasivkovou štěrbinu odkrytou a může jí bez odporu procházet vzduch k dýchání. Hlasové ústrojí tak využívá klidového postavení hlasivek. Při vytváření hlasu (fonaci) plní hlasivky jinou funkci – nacházejí se v tzv. hlasovém (fonačním) postavení. Výdechový proud vzduchu postupuje bez odporu z plic průdušnicí až k hrtanu. Zde se mu do cesty postaví překážka vytvořená hmotou hlasivek, které cestu vzduchu úplně uzavřou. Stažené hlasivky se pod tlakem vzduchu stávají pružnými a začínají kmitat. Hlasivky se přitom střídavě postupně otevírají a prudce uzavírají. V důsledku kmitání hlasivek se vzduchový proud „rozdrobí“ tak, že se víceméně pravidelně střídá vždy kvantum hustšího a řidšího vzduchu – vzniká tzv. vzduchová vlna, která je vnímána jako zvuk. Tento periodický proud vzduchových pulsů tvoří základ lidského hlasu. Bývá označován termínem základní (hlasivkový) tón a představuje nosný zvuk řeči. Frekvence kmitání hlasivek se označuje F_0 a nazývá se fundamentální frekvence nebo frekvence základního hlasivkového tónu. Tato frekvence je fyzikální charakteristikou řečového signálu a odpovídá takové výšce hlasu, jakou vnímá posluchač. Převrácená hodnota $T_0 = 1/F_0$ se nazývá periodou základního hlasivkového tónu. Frekvence kmitání je různá u dětí a dospělých, ale i u žen a mužů. Frekvence základního hlasivkového tónu však ve skutečnosti není konstantní (hlasivkové pulsy nejsou ryze periodické). V delších řečových úsecích se projeví vliv intonace promluvy. Délka periody i amplituda jednotlivých pulsů základního hlasivkového tónu se mírně liší v rámci velmi krátkého signálu. Takové mírné kolísání délky základní periody se nazývá jitter a je závislé na duševním stavu mluvčího. Kolísání amplitudy hlasivkových pulsů se označuje jako shimmer.

Klidové postavení hlasivek a postavení fonační představují dvě základní protikladné varianty činnosti hlasového ústrojí. Obou těchto poloh je využíváno při tvorbě řeči. Fonační postavení má za následek vznik hlasivkového tónu a používá se proto při vytváření znělých zvuků řeči. Neznělé zvuky jsou naopak tvořeny při klidovém postavení hlasivek, neobsahují tedy základní hlasivkový tón a vznikají až modifikací výdechového proudu vzduchu v nadhrtanových dutinách. Na rozdíl od prostého dýchání (když člověk mlčí) nebývají při vytváření neznělé řeči hlasivky úplně povoleny a ani hlasivková štěrbina nebývá úplně rozevřena [2].

1.1.3 Artikulační ústrojí

Artikulační ústrojí je posledním ústrojím, které se podílí na tvorbě řeči. Jeho význam spočívá v tom, že umožňuje vytvářet velké množství různých zvuků, které charakterizují mluvený jazyk. Skládá se jednak z nadhrtanových dutin a jednak z artikulačních orgánů, které jsou v těchto dutinách uloženy nebo je obklopují. Mezi nadhrtanové dutiny se řadí dutina hrdelní, ústní a nosní. Hranice mezi těmito dutinami tvoří čípek (uvula) a špička měkkého patra (velum), které zamezují nebo umožňují přístup vzduchu z dutiny hrdelní do dutiny nosní.

Zatímco se nadhrtanové dutiny účastní procesu tvorby řeči pasivně (nepohybují se), artikulační orgány (artikulátory) se účastní tvorby řeči většinou aktivně – tvoří pohyblivé součásti artikulačního ústrojí a svým pohybem mění velikosti nadhrtanových dutin. Z hlediska vytváření řeči mezi nejvýznamnější artikulátory řeči patří jazyk, rty a měkké patro, neboť se podílejí na vytváření největšího počtu různých zvuků. Dalšími artikulátory potom jsou zuby, tvrdé patro nebo čelisti. Artikulátorem je také hrtan, který kromě toho, že se zapojuje do tvorby znělosti, se také může pohybovat nahoru a dolů a měnit tak délku celého hlasového traktu.

Vůbec nejdůležitějším a nejsložitějším artikulátorem je jazyk. Skládá se ze tří částí: hrotu (špičky), hřbetu a kořene. Ty jsou součástí stejné svalové struktury jazyka, ale každá může do jisté míry fungovat nezávisle. Jazyk je tak velice pružný a přizpůsobivý, schopný tvořit mnoho tvarů a rychle přecházet z jedné pozice do druhé. Právě variabilita umístování jazyka vytváří nesčetné tvary ústní i hrdelní dutiny a vede tak k vytváření různých zvuků řeči.

Výrazných změn ve svém složení doznává zvuk v nadhrtanových dutinách, kam jako výdechový proud vzduchu (v podobě periodického proudu vzduchových pulsů nebo jako prostý proud vzduchu) postupuje z hlasového ústrojí. Tyto změny jsou v zásadě dvojího druhu:

- **Vytváření tónové struktury.** Při průchodu základního hlasivkového tónu nadhrtanovými dutinami dochází vlivem rezonance uvnitř hrdelní, ústní a popř. i nosní dutiny ke změně rozložení akustické energie ve vznikajícím řečovém signálu. Akustická energie se soustředí kolem určitých frekvencí, kterým říkáme formantové frekvence. Oblasti koncentrace (zesílení) akustické energie se nazývají formanty a označují se čísly, počínaje formantem s nejnižší frekvencí, F_1, F_2, \dots, F_n . Pokud se do procesu vytváření řeči zapojí i nosní dutina, dochází navíc vlivem jejích antirezonančních vlastností k potlačení některých frekvenčních oblastí, tzv. antiformantů (značí se A_1, A_2, \dots, A_m). Vzniká složený zvuk tónového charakteru, kterému říkáme hlas. Tvoří podstatu znělých částí řeči, především samohlásek. Různé zvuky (tj. zvuky s různou spektrální strukturou) přitom vznikají tak, že pohyblivé artikulátory mění tvar nadhrtanových dutin a tím i frekvenční vlastnosti vytvářených zvuků.
- **Vytváření šumové složky.** Artikulátory ovlivňují průchod vzduchu nadhrtanovými dutinami. Svým pohybem mohou průchod na různých místech zúžit, úplně uzavřít nebo naopak uvolnit. Výdechový proud vzduchu se pak prodírá přes vytvořené překážky a podle charakteru překážky vzniká šum různého druhu. Šum tvoří základ těch částí řeči, kterými při popisu jazyka říkáme souhlásky. Různé zvuky se v tomto případě vytvářejí pomocí různých typů překážek umístěných na různých místech hlasového traktu.

Během řeči dochází k plynulému pohybu všech artikulátorů. Lze si přitom představit, že po jistou velmi krátkou dobu každý artikulátor zůstává přibližně v určité jedné poloze a soubor všech artikulátorů tak tvoří jistou konfiguraci hlasového traktu.

Pohyby artikulátorů mění jednotlivé konfigurace podle předem daného programu. Řečník si totiž nejprve přeje vytvořit posloupnost zvuků, které odpovídají sdělení, jež chce vyjádřit, a potom mozek vysílá signály artikulátorům, jak a jakým způsobem mají fungovat. Výsledná posloupnost konfigurací hlasového traktu je tedy dána posloupností zvuků, které mluvčí vytváří. Je důležité si ale uvědomit, že ke změně konfigurací nedochází „najednou“ – hlasový trakt není diskretní systém. Přejednutí artikulátorů z jedné polohy do druhé tedy není okamžitý – dosažení cílového postavení řečových orgánů potřebuje jistý čas, který je závislý jednak na hmotnosti přemísťovaných orgánů a jednak na použitých svalových silách. Plynulost pohybu artikulačních orgánů při vytváření řeči má také za následek, že konfigurace hlasového traktu pro určitý zvuk nezávisí pouze na povaze tohoto zvuku, ale také na okolních zvucích v dané promluvě. Tento jev je typickým pro celý proces artikulace a nazývá se koartikulace. Projevuje se zejména při přechodu z jedné konfigurace hlasového traktu, dané právě vytvořeným zvukem do konfigurace jiné, dané následujícím zvukem, a to i při velmi pečlivé výslovnosti. Navenek se koartikulace projeví tím, že se akustické realizace stejného zvuku mění v závislosti na kontextu okolních zvuků [2].

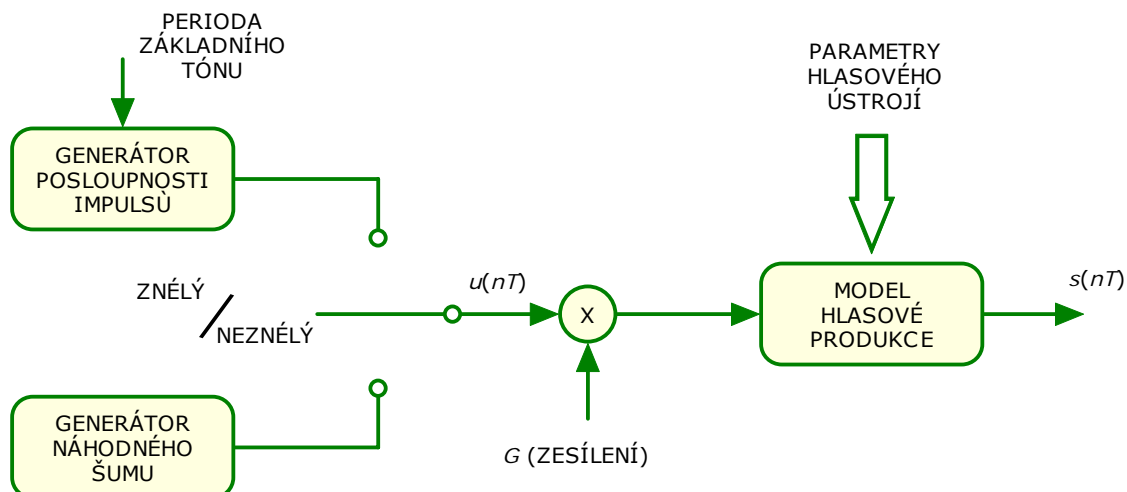
1.1.4 Modelování procesu vytváření řeči

Při počítačovém zpracování řeči bývá obvykle užitečné jistým způsobem modelovat procesy produkce i vnímání řeči a řeč následně reprezentovat pomocí parametrů zvoleného modelu. Cílem modelování produkce řeči je nalézt matematické vztahy, které by mohly být využity pro reprezentaci akustických fyzikálních dějů spojených s touto produkcí, a to s požadavkem minimální složitosti a maximální přesnosti modelu. Číselový model by měl respektovat především chování hlasivek, hlasového traktu a také proces vyzařování zvuku ze rtů, popř. z nosu. Model musí být přitom buzen odpovídajícím signálem podle znělosti generovaného zvuku.

Po sjednocení jednotlivých submodelů hlasivek, hlasového traktu a radiace a po zjednodušení výsledné přenosové funkce (viz [1] oddíl 2.2.2) vznikne přenosová funkce modelu nazývaného celopólový filtr

$$H(z) = \frac{1}{1 + \sum_{i=1}^Q a_i z^{-i}} \quad , \quad (1.1)$$

kde $Q = 2K+1$, K je počet formantů, které má model postihnout. Parametry modelu jsou sice proměnné v čase, ale na konkrétním časovém intervalu (10 – 30 ms) jsou uvažovány konstantní, jde o tzv. krátkodobý model. Na následujícím obrázku je uvedeno blokové schéma tohoto modelu [1].



Obr. 1.1 Model vytváření řeči [2].

2 Fonetika a logopedie

2.1.1 Fonetika

Fonetika (název z řeckého fóné – hlas, zvuk lidské řeči) je nauka o zvukové stránce lidské řeči a o její funkci při mluvení. Fonetika se zabývá tvořením hlásek, jejich vnímáním a užitím ve zvukové stavbě jazyka.

Nejnižšími stavebními jednotkami řeči jsou fonémy (hlásky), například [l], [e], [s]. Jejich kombinací se tvoří morfémy (nejmenší jazykové jednotky, které již mají význam), například [les], [lék], [mák], [hák]. Kombinací morfémů se tvoří slova. Slova dávají věty, věty promluvu apod. Fonetika si všímá nejen základních jednotek řeči (fonémy, hlásky), ale také vyšších zvukových jednotek (slabiky, slova) i jejich změn v souvislé řeči. Taktéž sleduje zvukové úpravy řeči různými modulačními faktory (melodie, tempo, přízvuk, barva hlasu).

Fonetika zkoumá zvuky lidské řeči z hlediska:

- jejich tvoření (artikulace),
- jejich akustických vlastností,
- jejich realizace v dorozumívacím procesu.

Dítě se ve svém vývoji musí naučit vybírat ze zvuků lidské řeči právě ty, které mají v systému mateřského jazyka fonologickou platnost, vyvíjí se jeho „fonematický sluch“. Tak dítě vycítí velmi brzy rozdíl mezi souhláskami znělými a neznělými (**ten-den**), později rozdíl mezi [t] a [k] (**tam-kam**) a mezi [l] a [r] (**líhá-rýha**), i když ke správné realizaci těchto fonémů dojde až později [3].

2.1.2 Logopedie

Logopedie je definována jako výchova řeči či nauka o výchově a vzdělání jedinců s vadami sdělování. Je součástí vývoje paradigmatu v oblasti speciální pedagogiky. Dominantním úkolem logopedické péče je korekce poruch řečové komunikace, proces, který je nerozlučně spojen s vývojem poznatků v oblastech medicínské a psychologické péče o osoby s vývojovými a získanými deficity a onemocněními.

Rozdělení vývojových poruch řečové komunikace:

Poruchy na bázi motorických řečových modalit:

- *dyslalie* (porucha realizace jednotlivých hlásek, vytváření jejich percepčně-motorických vzorů),
- *vývojová dyspraxie v řeči či artikulační neobratnost* (porucha v programování segmentů řeči, především hlásek),
- *palatolalie* (porušená realizace řeči při rozštěpu rtu a patra),
- *vývojová dysartrie* (porucha motoriky mluvy při organické lézi CNS),
- *dysfagie* (porucha polykání při poruše vývoje vitálních funkcí orofaciálního traktu),
- *rinolalie* (huhňavost, porucha nazality, změny rezonance a zvuku hlásek při mluvě),
- *dystonie* (porucha tvorby a užití hlasu, rezulující z poruch činnosti hlasivek či orofaciálního traktu),
- *balbuties* (kóktavost, porucha plynulosti mluvního projevu s tonickými či klonickými projevy),
- *tumultus sermonis* (breptavost, patologicky zvýšené tempo řeči se sníženou srozumitelností mluvního projevu).

Poruchy na bázi postižení vývoje individuálních jazykových schopností:

- opožděný vývoj řečových schopností, vývojová dysfázie, specifické poruchy učení.

Poruchy na bázi postižení kognitivních schopností v řečové komunikaci – kognitivně komunikační poruchy (KKP):

- KKP při pervazivním vývojovém onemocnění, KKP u stavů po úrazu CNS či onemocnění CNS, KKP při mentálním handicapu.

Poruchy na bázi postižení percepce:

- poruchy řečové komunikace vyvolané vlivy poruchy sluchu, poruchy řečové komunikace vyvolané vlivy poruchy zraku [4].

Dyslalie:

Dyslalie je nejrozšířenější odchylkou ve vývoji řečových schopností, vznikající nejčastěji v průběhu předškolního věku. Zahrnuje artikulaci jedné či více hlásek způsobem motoricky a především zvukově nápadným a odchýlným od kodifikované a uznávané výslovnosti.

Tyto odchylky bývají v odborné literatuře často označovány řeckými písmeny s koncovkou a označením místa či způsobu odchylné artikulace. Nejčastějšími odchylkami jsou sigmatismy (odchylná artikulace sykavek) a rotacismy (odchylná artikulace vibrant – hlásek [r] a [ř]). Frekventovaným závažným mechanismem je verální rotacismus, tedy tvorba hlásky [r] či [ř] pomocí vibrace v hrdelní oblasti nejčastěji s pomocí okraje měkkého patra (ráčkování), a interdentalní sigmatismus (mezizubní poloha hrotu jazyka při artikulaci sykavky), případně addentalní sigmatismus (hrot jazyka se dotýká alveol zubů), kdy je tvořen zvuk nazývaný v obecné mluvě „šišlání“. Právě přítomnost „lidových“ výrazů pro tyto nejčastější odchylky artikulační normy vypovídá o jejich frekventovanosti, nápadnosti a vlivu na sociální stigmatizaci řečového projevu.

Do oblasti dyslalií bývají řazeny pojmy mogilalie a paralalie. Ty zahrnují povětšinou přechodné vývojové jevy, při nichž dítě danou hlásku vynechává (mogilalie) či zaměňuje za jinou hlásku (paralalie). Z vývojového hlediska není vhodné tyto jevy řadit k patologickým, jsou součástí vývoje řečových schopností u všech dětí.

Vytvořené a zafixované odchylky v artikulaci hlásek přetrvávají po celý život, pokud nejsou upraveny specializovanou logopedickou péčí. Tyto metodiky má mít pod kontrolou zkušený klinický či školský logoped. Při jejich neodborném užití hrozí nebezpečí neúspěchu a následně fixování názorů o neúspěšnosti či extrémní náročnosti v odstraňování poruch artikulace. Úprava artikulačních vzorů hlásek je realizovatelná i u dospělých osob, kde lze výborně využít a posunout jejich stupeň sluchové percepce a motoriky, ovšem pouze v případě aktivní dlouhodobé motivace [4].

Zásady logopedické diagnostiky:

Logopedická diagnostika má za úkol zachytit celkový stav a poruchy komunikace řečí, jazykovými prostředky, písmem a s nimi spojené deficity praxe, gnosie a kognitivních funkcí.

Jejími základními cíli jsou:

- zjištění přítomnosti poruchy řečové komunikace, její tíže a vážnosti,
- diferenciální diagnostika poruch řečové komunikace,
- zpracování návrhu následné logopedické intervence [4].

Všeobecné směrnice pro úpravu dyslalií:

Jednotlivé hlásky z fonetického hlediska nejsou univerzálními ustálenými zvuky, ale proměňují se (výškou, trváním, silou) podle svého postavení a kontextu s ostatními hláskami. Je tak nezbytné akceptovat alespoň hlásku v pozici iniciální (na začátku slova), mediální (uprostřed slova) a finální (na konci slova). Například obávanou hlásku [r] je třeba do řeči zabudovat nejméně v deseti spojeních. Je známou skutečností, že po vyvození hlásky [r] v preferované (trubce) nebo (trnce) děti vyslovují finální [r] velice přehnaným způsobem (svetrrr, Petrrr), protože se dosud procvičením nedomohly fonetického povědomí, že hláska [r] ve finálním postavení je asi o půl kmitu kratší.

Prvním předpokladem pro úspěšnou úpravu dyslalie je vyšetření sluchu. A to sluchu jako schopnosti slyšet, ale také sluchu fonemického, jako schopnosti diferencovat zvuky lidské řeči. Doporučuje se rovněž ověření úrovně muzických elementů řeči dítěte, které jsou nadřazeny artikulační schopnosti. Důvodem je, že dospělý člověk v určitém časovém odstupu začíná rozumět patlavé řeči, pokud dítě komolená slova pronáší se správným přízvukem, s odpovídajícími prozodickými faktory.

Při úpravě dyslalie se postupuje vždy přísně individuálně. Prioritní je pro dítě i logopeda pocit úspěšnosti. Začíná se tedy hláskou, jejíž vyvození má reálnou naději na úspěch. Vzbudí se tak radost v dítěti i jeho okolí, posílí se motivace. Rozumě se spolupracuje s rodiči jako s partnery. Doma by měli rodiče podle pokynů procvičovat pouze ta slova a spojení, která se dítěti v průběhu terapeutického setkání dařila. Podnítí se tím nezbytná fáze fixace vyvozené hlásky. Efektivní je kopie souboru procvičovaných spojení z logopedické poradny, kde se označí úspěšná slova [5].

3 Analýza řečového signálu

Základem většiny metod analýzy akustického signálu řeči je předpoklad, že se jeho vlastnosti mění pomalu. Tento předpoklad vychází z kapitoly 1.1 a vede na aplikaci tzv. metod krátkodobé analýzy, při nichž se úseky řečového signálu vydělují a zpracovávají tak, jako by to byly oddělené krátké zvuky, tzv. segmenty. Signál je pak v každém segmentu přibližně stacionární. To znamená, že má signál střední hodnotu a rozptyl přibližně konstantní. Délka segmentu se volí 10 – 20 ms. Pro extrakci příznaků z řečového signálu byly použity metody PLP a MFCC, které jsou dále popsány [2].

3.1 Lineární prediktivní (předpovědní) analýza

Lineární prediktivní analýza je jednou z nejefektivnějších metod analýzy akustického signálu. Je to metoda, která se snaží na krátkodobém základu odhadnout přímo z řečového signálu parametry modelu vytváření řeči, jehož struktura byla odvozena v odstavci 1.1.4. Atraktivnost metody spočívá v její schopnosti zabezpečit odhad uvedených parametrů v relativně přijatelné výpočetní zátěži. Odvození koeficientů lineární predikce vychází z modelu vytváření řeči, který se skládá ze systému s časově proměnným přenosem a generátoru budících funkcí, který budí systém při vytváření znělých zvuků posloupností impulsů a při vytváření neznělých zvuků náhodným šumem. Princip metody lineární predikce je tedy založen na předpokladu, že k -tý vzorek signálu $s[n]$ lze popsat lineární kombinací Q předchozích vzorků a buzení $u[n]$, tj.

$$s[n] = -\sum_{i=1}^Q a_i s[n-i] + Gu[n], \quad (3.1)$$

kde G je koeficient zesílení a Q je řád modelu. Koeficienty lineární predikce pak odpovídají zpětně odvozeným koeficientům lineárního číslicového filtru obsaženého v modelu vytváření řeči. Proces modelování lze popsat přenosovou funkcí

(3.2) a Obr. 1.1.

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{A(z)} = \frac{G}{1 + \sum_{i=1}^Q a_i z^{-i}} \quad (3.2)$$

Sledovanými parametry jsou zde koeficienty a_i číslicového filtru a koeficient zesílení G . Je-li splněn předpoklad přibližné stacionarity signálu na sledovaném časovém intervalu, lze pro určení a_i a G využít autokorelační metodu (viz [1] odstavec 4.5.1.), kde se koeficienty a_i určují řešením normálních rovnic pomocí autokorelační posloupnosti. Vzniklá autokorelační matice je Toeplitzovou maticí, která je řešena Levinsonovým – Durbinovým algoritmem.

Tento účinný popis spektrálních vlastností řečového signálu však nekoresponduje příliš dobře se způsobem, jak slyší zvuky člověk. Zmírnění tohoto nedostatku přináší systematický a velmi účinný postup nazvaný Perceptivní lineární prediktivní analýza,

který byl navržen Hermanským (viz. [6]) a který je podrobněji popsán v následujícím odstavci [2].

3.1.1 Perceptivní lineární prediktivní analýza (PLP)

Pro transformaci výkonového spektra řečového signálu do odpovídajícího sluchového spektra kombinuje tato metoda tři složky z psychofyziky slyšení: kritické pásmo spektrální citlivosti, křivky stejné hlasitosti a vztah vyjadřující závislost mezi intenzitou zvuku a jeho vnímanou hlasitostí. Sluchové spektrum je pak aproximováno autoregresním celopólovým modelem podobně jako u standardní metody lineární prediktivní analýzy. V následujících odstavcích jsou ve stručnosti popsány jednotlivé kroky PLP analýzy [2].

- **Výpočet výkonového spektra řečového signálu.** Řečový signál je nejprve segmentován a poté vážen Hamingovým okénkem. Z daného segmentu $s[n]$ jsou pomocí FFT vypočítány vzorky signálového spektra $S(\omega)$. Krátkodobé výkonové spektrum $P(\omega)$ je pak definováno vztahem

$$P(\omega) = |S(\omega)|^2 = [\operatorname{Re} S(\omega)]^2 + [\operatorname{Im} S(\omega)]^2. \quad (3.3)$$

- **Nelineární transformace frekvencí a kritická pásma spektrální citlivosti slyšení.** Člověk nevnímá změny ve výšce zvuku lineárně, ale spíše logaritmicky. Vnímání zvuků je též ovlivněno tzv. maskováním zvuků. Modelování těchto jevů je v PLP analýze realizováno jednak nelineární transformací původní osy frekvencí ω [rad/s] na osu frekvencí $\Omega(\omega)$ měřenou v jednotce bark, a to podle vztahu

$$\Omega(\omega) = 6 \ln \left(\frac{\omega}{1200\pi} + \sqrt{\left(\frac{\omega}{1200\pi} \right)^2 + 1} \right), \quad (3.4)$$

kde $\omega = 2\pi f$ [rad/s] a $\Omega(\omega)$ [bark], a dále konstrukcí maskujících křivek (v podstatě jde o pásmové propusti), které simulují kritická pásma slyšení. Základní charakteristiku pásmové propusti lze popsat vztahem

$$\Psi(z) = \begin{cases} 0 & \text{pro } z < -2,5 \\ 10^{z+0,5} & \text{pro } -2,5 \leq z \leq -0,5 \\ 1 & \text{pro } -0,5 < z < 0,5 \\ 10^{-2,5(z-0,5)} & \text{pro } 0,5 \leq z \leq 1,3 \\ 0 & \text{pro } z > 1,3 \end{cases}, \quad (3.5)$$

kde z je v jednotkách [bark]. Amplitudová frekvenční charakteristika tohoto filtru $\Psi(z)$ je asymetrická a má strmost +20 dB/bark směrem k nižším frekvencím (pro $z \leq -0,5$) a -50 dB/bark směrem k vyšším frekvencím (pro $z \geq 0,5$). Filtry jsou ve zpracovávaném frekvenčním pásmu rozmístěny lineárně s krokem přibližně 1 bark. Nultý filtr má obvykle „střed“ v počáteční

hodnotě přenášeného pásma (tedy 0 bark) a poslední $M-1$ filtr v horní hodnotě frekvence přenášeného pásma (pro přenášené pásmo 0 až 4000 Hz je to 15,57 bark). Příspěvky obou krajních filtrů se obvykle nepočítají. Pro výpočet predikce (viz. [2]) se určí jejich hodnoty tak, že se položí rovny hodnotám pravého resp. levého souseda.

- **Přizpůsobení kritických pásmových filtrů křivkám stejné hlasitosti.** Člověk vnímá intenzitu zvuku v závislosti na frekvenci jako hlasitost. Abychom mohli přizpůsobit výkonové spektrum $P(\omega)$ této vlastnosti lidského sluchu, provedeme v dalším kroku nejprve preemfázi diskretních vzorků křivek simulující pásmový filtr m -tého kritického pásma a odpovídajících hodnot (vzorků) aproximující křivky $E(\omega)$

$$\Phi_m(\Omega(\omega)) = E(\omega)\Psi(\Omega(\omega) - \Omega_m), \quad (3.6)$$

nebo pro frekvence Ω

$$\Phi_m(\Omega) = E(\Omega)\Psi(\Omega - \Omega_m), \quad (3.7)$$

kde Ω_m [bark] je „střední“ frekvence m -tého kritického pásmového filtru $m = 0, \dots, M-1$. $E(\omega)$ resp. $E(\Omega)$ je funkce navržena jako přibližná aproximace na stejnou citlivost lidského sluchu v odlišných frekvencích.

- **Vážená spektrální sumarizace vzorků výkonového spektra.** Jsou-li vyjádřeny hodnoty výkonového spektra $P(\omega)$, lze „průchod“ tohoto spektra m -tým kritickým pásmovým filtrem $\Phi_m[\Omega]$, jež je přizpůsoben křivkám stejné hlasitosti, vyjádřit jako

$$\Theta(\Omega_m) = \sum_{\Omega=\Omega_m-2,5}^{\Omega_m+1,3} P(\Omega)\Phi_m(\Omega) = \sum_{\omega=\omega_{md}}^{\omega_{mh}} P(\omega)\Phi(\Omega(\omega)), \quad (3.8)$$

kde $m = 1, \dots, M-2$. V praxi je vhodné hodnoty $\Theta(\Omega_m)$ počítat pro diskretní proměnnou ω (využívanou například v FFT), a to z přepočtených hodnot $\Phi_m[\Omega(\omega)]$ a vždy aktuálních hodnot $P(\omega)$. Sumační meze ω_{md} a ω_{mh} lze určit na základě vztahu inverzního k (3.4), tedy jako $\omega_{md}=1200\pi\sinh((\Omega_m-2,5)/6)$, resp. $\omega_{mh}=1200\pi\sinh((\Omega_m+2,5)/6)$.

- **Uplatnění vztahu vyjadřujícího závislost mezi intenzitou zvuku a vnímanou hlasitostí.** Hodnoty $\Theta(\Omega_m)$ jsou v následujícím kroku podrobeny operaci, jež je aproximací zákona „slyšení“ vyjadřujícího vztah mezi intenzitou zvuku a jeho vnímanou hlasitostí (viz. [2] oddíl 3.2)

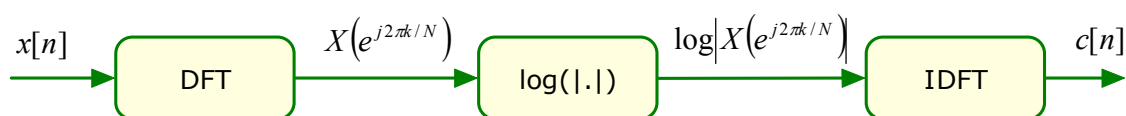
$$\xi(\Omega_m) = (\Theta(\Omega_m))^{0,3}, \quad (3.9)$$

kde $m = 1, \dots, M-2$. Tato operace též zajišťuje redukci amplitudové proměnlivosti „výstupů“ kritických pásmových filtrů, což má za následek, že hledaný selopólový model může být relativně nízkého řádu.

- **Aproximace spektrem celopólového modelu.** V dalším kroku PLP analýzy jsou hodnoty $\xi(\Omega_m)$, $m = 1, \dots, M-2$, aproximovány spektrem celopólového modelu. To znamená, že jsou určeny hledané koeficienty lineární predikce (viz. úvod této kapitoly) [2] oddíl 4.5.1 a 4.5.2.

3.2 Homomorfní zpracování řeči

Homomorfní analýza patří ke skupině postupů nelineárního zpracování signálů, které jsou založeny na využití zobecněného principu superpozice. Tyto postupy se hodí pro analýzu (přesněji oddělování) signálů, které vznikly konvolucí či vynásobením dvou nebo více složek. Vhodnost aplikace takového přístupu při zpracování řeči vyplývá z dříve uvedeného modelu generování řečového signálu konvolucí budící funkce a impulsní odezvy hlasového ústrojí. Cílem analýzy je určit parametry systému, tj. vydělit jednotlivé členy konvolutorního součinu. Metoda vydělení členů konvolutorního součinu, založená na teorii homomorfních systémů vzhledem ke konvoluci, se nazývá „homomorfní filtrace“. Základní vlastností homomorfních systémů je tzv. kepstrum signálu. Postup výpočtu kepstra $c[n]$ ze signálu $x[n]$ je naznačen na obr.3.2. Mezi metody homomorfní filtrace patří také metoda melovských kepstrálních koeficientů, která byla použita v této práci[2].



Obr. 3.1 Postup výpočtu kepstra pomocí DFT [2].

3.2.1 Melovské kepstrální koeficienty (MFCC)

Podobně jako PLP je i zpracování pomocí MFCC navrženo tak, aby do jisté míry respektovalo nelineární vlastnosti vnímání zvuků lidským uchem. MFCC se snaží „kompenzovat“ zejména nelineární vnímání frekvencí, a to využitím banky trojúhelníkových pásmových filtrů lineárním rozložením frekvencí v tzv. melovské frekvenční škále (viz [2] oddíl 3.2), jež je definována vztahem

$$f_m = 2595 \log_{10} \left(1 + \frac{f}{700} \right), \quad (3.10)$$

kde f [Hz] je frekvence v lineární škále a f_m [mel] je odpovídající frekvence v nelineární melovské škále. Proces určení melovských kepstrálních koeficientů lze popsat následujícím postupem.

Na vstup systému jsou přiváděny vzorky řečového signálu $s[n]$. Je provedena preemfáze signálu a na segmenty signálu délky obvykle 10 až 30 ms je aplikováno nejčastěji Hammingovo okénko. Přesná časová délka okénka (počet vzorků při dané frekvenci vzorkování F_{vz}) se volí rovna mocnině 2 vzhledem k následnému zpracování

FFT. Přitom je doporučováno posouvat okénko o časový úsek 10 ms, tj. výsledná parametrizace řečového signálu je vyčíslována 100krát za sekundu.

V dalším bloku zpracování se pomocí FFT vypočte modul spektra $|S(f)|$ analyzovaného signálu případně výkonové spektrum $|S(f)|^2$.

Klíčová část celého procesu zpracování je melovská filtrace. Výpočetní algoritmus této filtrace je realizován bankou trojúhelníkových pásmových filtrů podél frekvenční osy s měřítkem v melovské škále. Počet pásem této banky filtrů je rozumné volit v závislosti na počtu a umístění kritických pásem, a to při respektování velikosti vzorkovací frekvence F_{vz} [Hz] a celkové šířky přenášeného pásma B_w [Hz], resp. B_{mw} [mel]. Trojúhelníkové filtry jsou standardně rozloženy přes celé frekvenční pásmo od nuly až do Nyquistovy frekvence. Pro střední frekvence jednotlivých filtrů $b_{m,i}$ platí v melovské škále vztah

$$b_{m,i} = b_{m,i-1} + \Delta_m, \quad (3.11)$$

kde $b_{m,0} = 0$ mel, $i = 1, 2, \dots, M^*$, a $\Delta_m = B_{mw} / (M^* + 1)$. Odezvy jednotlivých filtrů mají v melovské frekvenční škále tvar rovnoramenných trojúhelníků a jsou rovnoměrně rozloženy ve frekvenci. Pro výpočet odezev filtrů se však musí přepočítat všechny koeficienty FFT do melovské frekvenční škály. Alternativním postupem, který se spíše využívá, je vyjádření trojúhelníkových filtrů ve frekvenční škále s měřítkem v hercích při současném využití původních koeficientů získaných FFT. V tomto případě pak další postup spočívá v přepočtu všech středních frekvencí $b_{m,i}$, $i = 1, \dots, M^* + 1$, s využitím inverzního vztahu k (3.10) (tj. vztahu $f = 700[\exp(0,887 \cdot 10^{-3} f_m) - 1]$) na střední frekvence b_i , $i = 1, \dots, M^* + 1$, vyjádřené v jednotce [Hz]. Odezvy filtrů lze pak vyjádřit vztahem

$$y_m(i) = \sum_{f=b_{i-1}}^{b_{i+1}} |S(f)| u(f, i) \quad i = 1, 2, \dots, M^*, \quad (3.12)$$

kde frekvence f jsou vybírány ze souboru frekvencí využívaných při výpočtu FFT a $u(f, i)$ je vyjádření trojúhelníkového filtru, který lze popsat vztahem

$$u(f, i) = \begin{cases} \frac{1}{b_i - b_{i-1}} (f - b_{i-1}) & \text{pro } b_{i-1} \leq f < b_i \\ \frac{1}{b_i - b_{i+1}} (f - b_{i+1}) & \text{pro } b_i \leq f < b_{i+1} \\ 0 & \text{pro ostatní případy.} \end{cases} \quad (3.13)$$

Průchod signálu filtrem tedy znamená, že každý koeficient FFT je násoben odpovídajícím ziskem filtru a výsledky jsou pro příslušné filtry akumulovány.

Další krok spočívá ve výpočtu logaritmu výstupů $y_m(i)$ jednotlivých filtrů. Logaritmováním akumulovaných koeficientů se vkládá do procesu zpracování důležitý příznak keprstrální analýzy, který jednak příznivě omezí dynamiku signálu (podobně

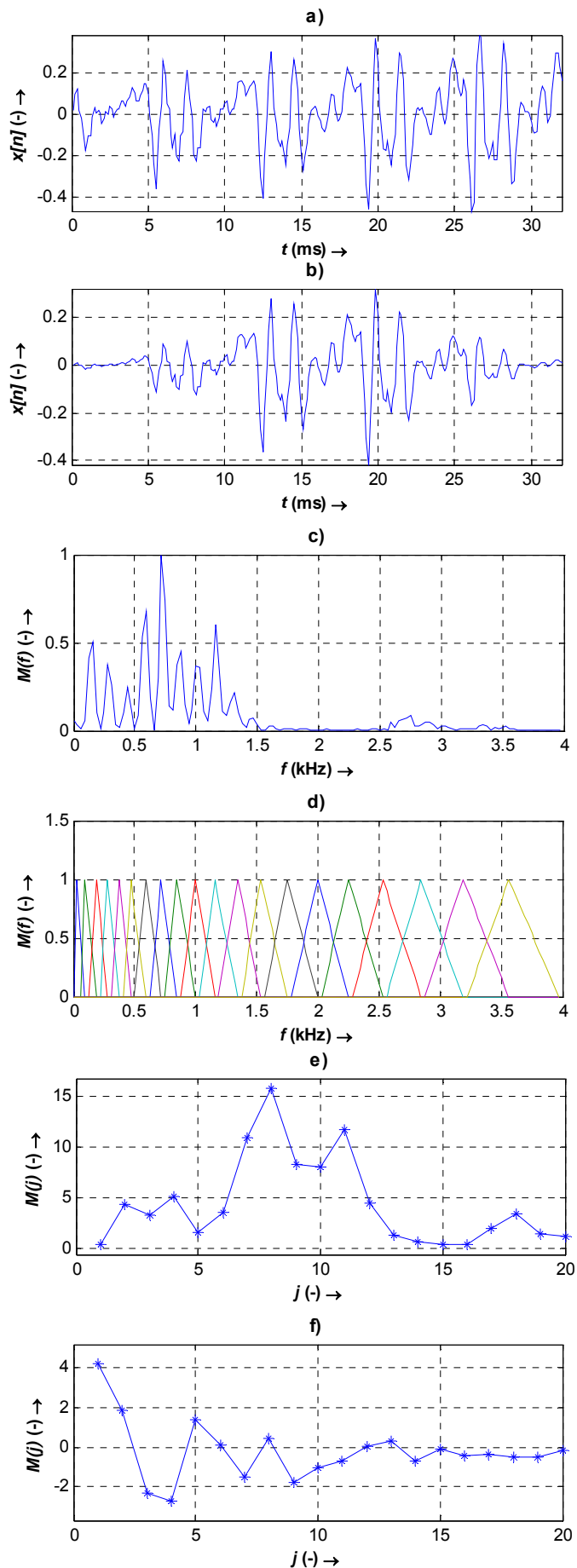
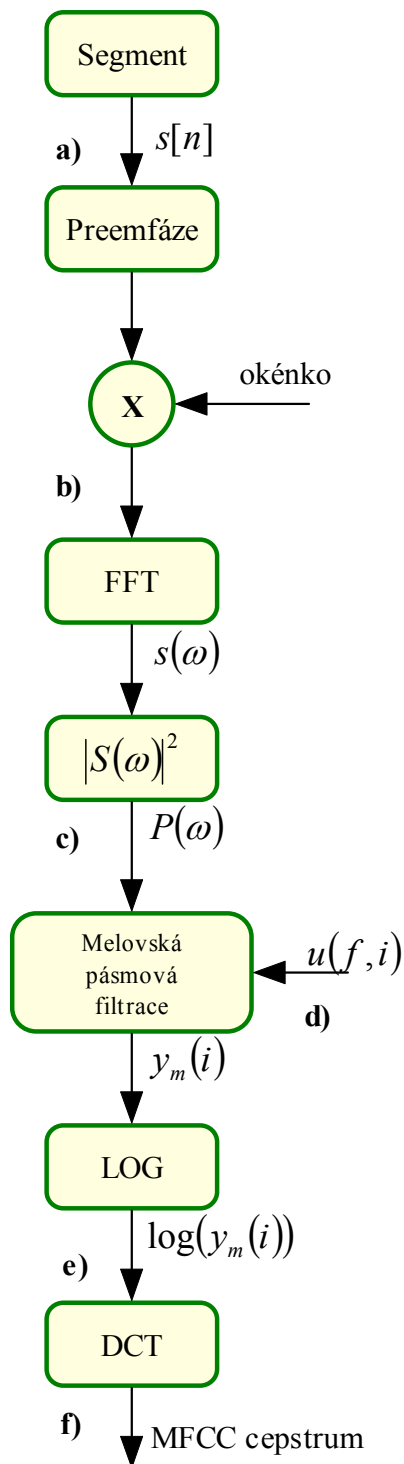
jako to dělá i lidské ucho). Posledním krokem při výpočtu melovských kepstrálních koeficientů $\{c_m(j)\}_{j=1}^M$ je provedení IFFT. Vzhledem k tomu, že výkonové spektrum je reálné a symetrické, bude se IFFT redukovat na DCT

$$c_m(j) = \sum_{i=1}^{M^*} \log y_m(i) \cos\left(\frac{\pi j}{M^*}(i-0,5)\right), \text{ pro } j=0,1,\dots,M, \quad (3.14)$$

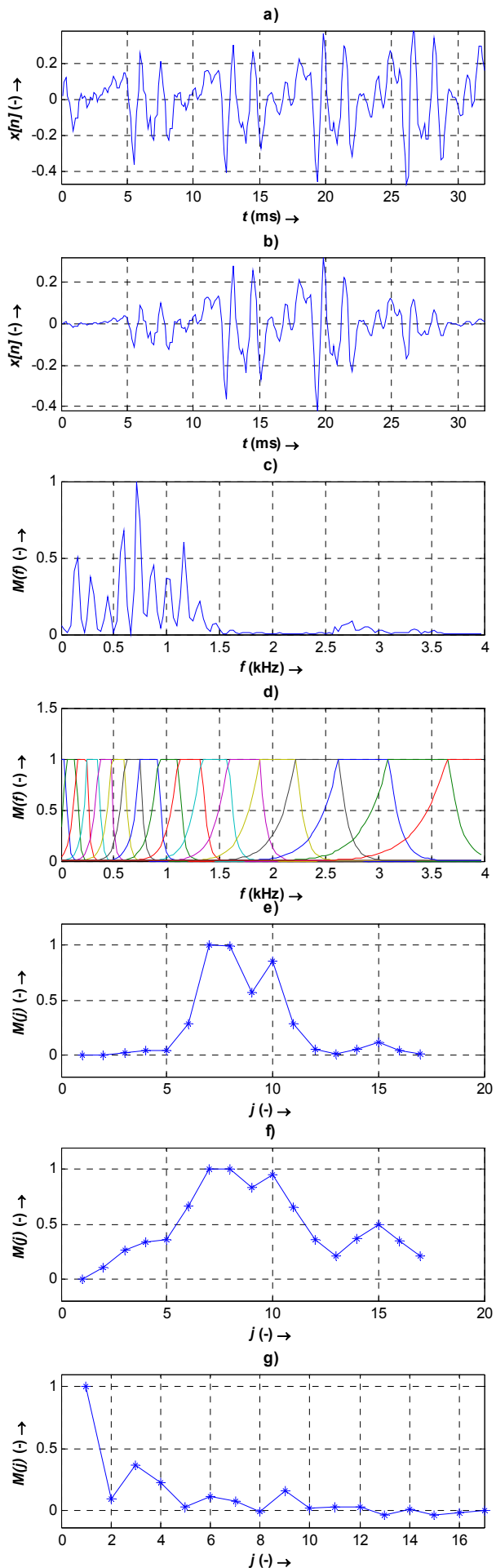
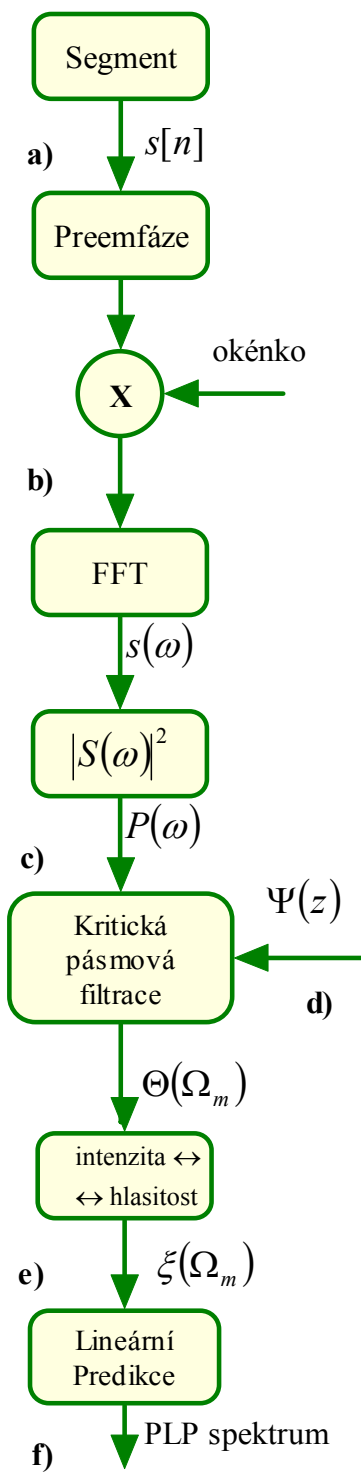
kde M^* je počet pásem melovského pásmového filtru a M je počet melovských kepstrálních koeficientů [2].

Princip metody MFCC je naznačen v Obr. 3.3. Průběhy a), b), c), e) a f) jsou příkladem stejně označených výstupních signálů příslušných bloků zpracování, průběh d) je příklad melbanky.

Metodu PLP popisuje Obr. 3.4., kde průběhy a), b), e) a f) jsou příkladem stejně označených výstupních signálů příslušných bloků zpracování, průběh d) je příklad melbanky.



Obr. 3.2 Blokové schéma MFCC.



Obr. 3.3 Princip PLP.

4 Segmentace slov na hlásky

Klasifikace příznaků do tříd je prováděna na úrovni hlásek. Z toho důvodu je nutné před vlastní klasifikací nahrané slovo segmentovat na jednotlivé hlásky. Segmentace slov na hlásky může být provedena různými způsoby. V této práci k tomu byla použita metoda sledování rozdílnosti příznaků.

4.1 Metoda sledování rozdílnosti příznaků

Metoda sledování rozdílnosti příznaků vychází z toho, že řečový signál má kvazistacionární charakter. To znamená, že se předpokládá, že akustický signál má po celou dobu trvání hlásky stejný charakter. Při změně vyslovované hlásky se změní i charakter akustického signálu, ale signál po celou dobu trvání jedné hlásky bude mít opět kvazistacionární charakter. Tento závěr vychází z toho, že artikulační ústrojí má omezenou rychlost změny parametrů (tj. charakteru generovaného signálu). Charakter signálu v nějakém časovém intervalu vystihují příznaky ze segmentu řečového signálu.

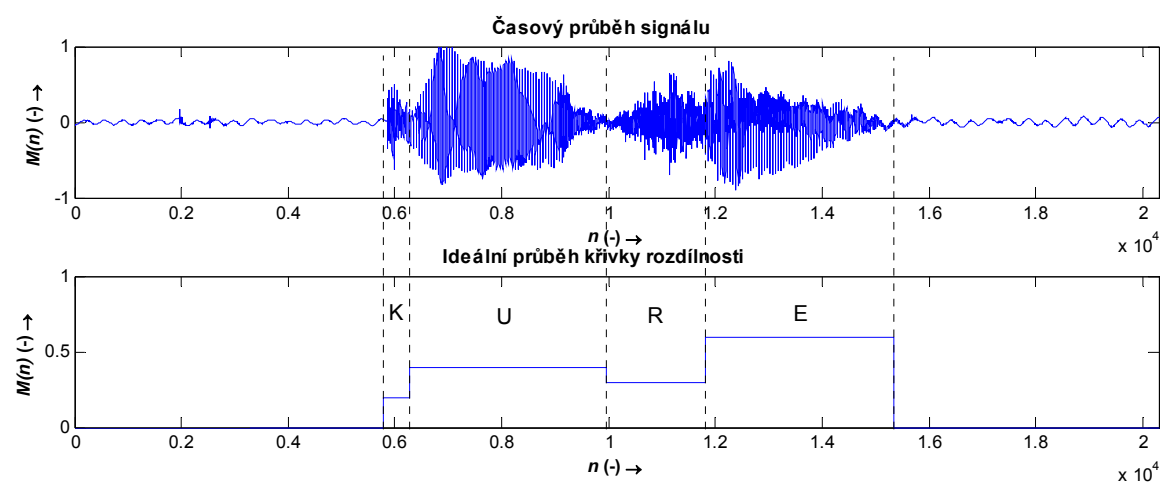
Pro extrakci příznaků mohou být použity například metody pro výpočet energie, počtu průchodu signálu nulou nebo zmíněné MFCC a PLP koeficienty. Podle článku [11] se jako vhodné příznaky jeví spektrální koeficienty filtrované bankou melovských filtrů. Jde v podstatě o metodu MFCC, která je popsána v odstavci 3.2.1, kde se po melovské filtraci již nepočítá kepsrum.

Pro příznaky jednotlivých segmentů řečového signálu se dále vypočítá křivka rozdílnosti příznaků. Úkolem křivky rozdílnosti příznaků je zachytit změny v časovém průběhu jednotlivých příznaků podél doby trvání celého akustického signálu. Algoritmus výpočtu křivky musí být takový, aby křivka nabývala malých hodnot v oblastech, kde se nacházejí jednotlivé fonémy (příznaky zde téměř nemění svoji hodnotu) a naopak hodnot velkých v oblastech přechodů mezi fonémy (příznaky se zde mění z jedné konstantní hodnoty na druhou). Ideální průběh křivky je uveden v 0.1. Tato křivka je získána tak, že se sledují absolutní změny okamžitých hodnot daných příznaků. Křivka rozdílnosti příznaků je definována následujícím vztahem

$$B(j) = \sum_k^{K-1} \left(\left| P_k(j) - P_k(j+1) \right| \right), \quad (4.1)$$

kde $P_k(j)$ je hodnota sledovaného příznaku k v segmentu j a K je počet příznaků v jednom segmentu.

Výsledná křivka $B(j)$ ve většině případů obsahuje velké množství maxim. Proto se křivka dále prahuje experimentálně zvoleným prahem. Všechna maxima pod zvoleným prahem se vynulují, maxima nad prahem odpovídají výrazným změnám příznaků, které korespondují se změnou hlásky [10].



Obr. 4.1 Princip segmentace slova na hlásky.

5 Klasifikace příznaků řečového signálu

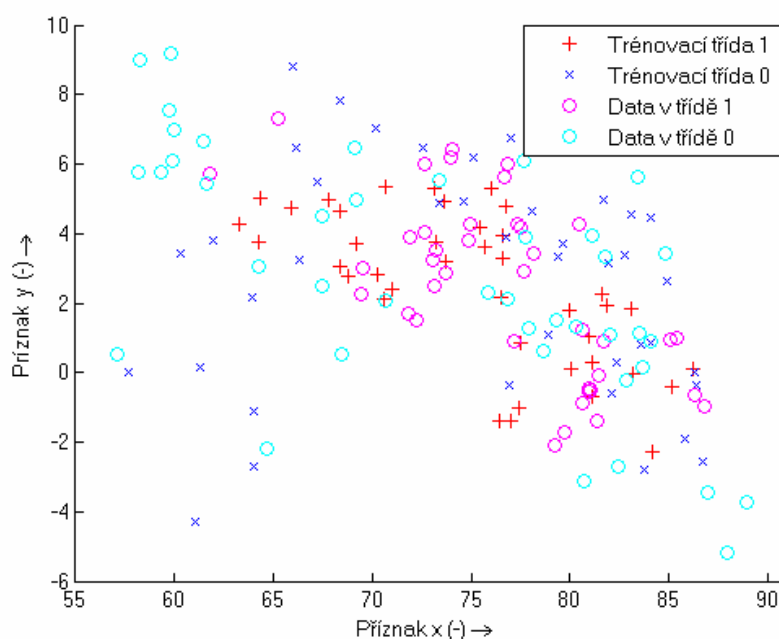
Klasifikace obecně znamená třídění, hodnocení nebo posuzování. V případě klasifikace příznaků řečového signálu znamená klasifikace zařazení extrahovaných příznaků do určitých tříd. V případě rozpoznání logopedických vad jde o klasifikaci do dvou tříd, a to do třídy správné výslovnosti nebo vadné výslovnosti. V dnešní době existuje několik klasifikátorů založených na různých algoritmech od jednoduchých až po složité, které vyžadují strojové učení. Tyto metody jsou také nazývány jako metody rozpoznání vzoru. V této práci budou pro klasifikaci příznaků řečového signálu testovány klasifikátory KNN, SVM, GMM a ANN, které jsou podrobněji popsány níže.

5.1 Algoritmus k -nejbližších sousedů (k -NN)

Klasifikátor k -NN přiřadí testovacím příznakům tu třídu, do které spadá většina k nejbližších sousedů z dat trénovacích.

V první fázi (učení) se každý trénovací bod o N příznacích umístí do odpovídajícího místa v N -rozměrném prostoru. Každý trénovací bod má pevně definovanou svojí třídu. V druhé fázi (klasifikace) se klasifikovaný bod o N příznacích umístí do téhož prostoru. Dále je na základě Euklidovské nebo jiné metriky vyhledáno k sousedních (nejbližších) bodů a podle převládající třídy těchto bodů je určena třída bodu klasifikovaného. Příklad klasifikace pro $k = 2$ ukazuje následující obrázek. Aby bylo možné body zobrazit v dvourozměrné rovině, jsou každému bodu přiděleny pouze dva příznaky.

Trénovací data jsou v obrázku vyznačeny křížkem a klasifikovaná data kolečkem, jednotlivé třídy jsou barevně odlišeny.

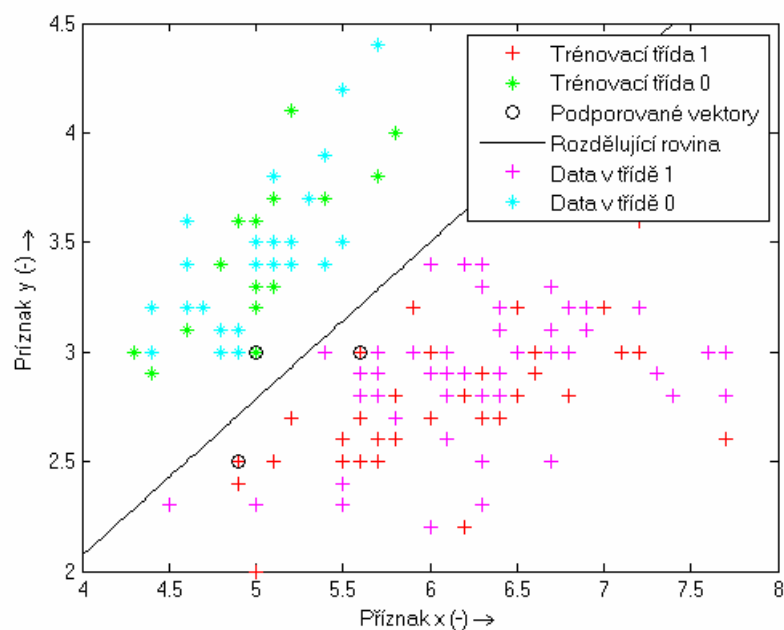


Obr. 5.1 Příklad k -NN klasifikace.

5.2 Algoritmy podpůrného učení (SVM)

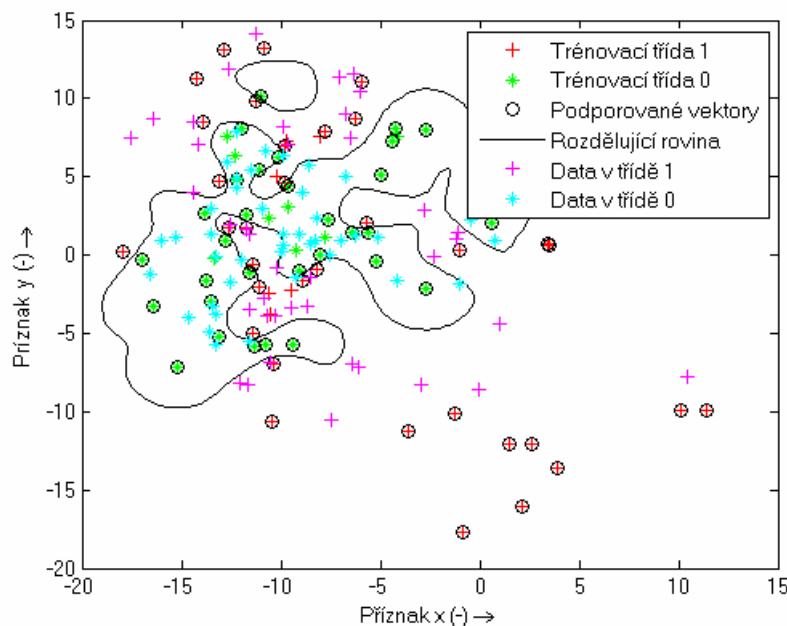
Klasifikátor SVM třídí data na základě tak zvaných podporovaných vektorů, proto název Support Vector Machine.

Stejně jako u k -NN se v první fázi (učení) každý trénovací bod o N příznacích umístí do odpovídajícího místa v N -rozměrném prostoru. Každý trénovací bod má pevně definovanou svoji třídu. Algoritmus dále hledá rovinu (v dvourozměrné rovině přímku viz obr) pomocí podporovaných vektorů, která rozdělí data do tříd. V druhé fázi (klasifikace) se klasifikovaný bod o N příznacích umístí do rozděleného prostoru a podle umístění je určena jeho třída.



Obr. 5.2 Příklad SVM klasifikace použitím lineární funkce.

Ne vždy je však rozdělení prostoru tak jednoduché jako v Obr. 5.2. Na Obr. 5.3 již není možné rovinu rozdělit lineárně. Proto SVM pro rozdělení prostoru využije místo lineární jádrové funkce jádrovou funkci RBF.



Obr. 5.3 Příklad SVM klasifikace použitím funkce RBF.

5.3 Smíšené Gaussovy modely (GMM)

GMM patří mezi metody využívající statistické rozpoznání vzorů. Tyto metody vycházejí z předpokladu, že modely stejných tříd mohou mít některé podobné statistické vlastnosti. GMM modeluje třídy příznaků pomocí směsí Gaussových funkcí.

Rozdělení pravděpodobnosti příznakových vektorů se může popsat směsí Gaussových hustotních funkcí, to znamená váženou lineární kombinací hustotních funkcí normálních rozdělení jednotlivých tříd ve tvaru

$$p(o | \lambda^s) = \sum_{i=1}^{M^s} w_i^s p_i^s(o). \quad (5.1)$$

Hodnota M^s udává počet Gaussových funkcí rozložení pravděpodobnosti s , $w_i^s, i=1, \dots, M^s$, jsou váhy jednotlivých složek vyhovující podmínce

$$\sum_{i=1}^M w_i^s = 1 \quad (5.2)$$

a $p_i^s(o), i=1, \dots, M^s$, značí hustoty pravděpodobnosti jednotlivých složek reprezentované n -rozměrnou normální hustotu pravděpodobnosti se střední hodnotou μ_i^s a kovarianční maticí C_i^s , tj.

$$p_i^s(o) = \frac{1}{(2\pi)^{n/2} |C_i^s|^{1/2}} \exp\left(-\frac{1}{2} (o - \mu_i^s)^T (C_i^s)^{-1} (o - \mu_i^s)\right), \quad (5.3)$$

kde n je dimenze příznakových vektorů. Protože směs Gaussových hustotních funkcí je zcela určena vektory středních hodnot, kovariančními maticemi a váhami

jednotlivých složek, lze model třídy s reprezentovaný touto směsí charakterizovat zápisem

$$\lambda^s = \{w_i^s, \mu_i^s, C_i^s\}, i = 1, \dots, M^s. \quad (5.4)$$

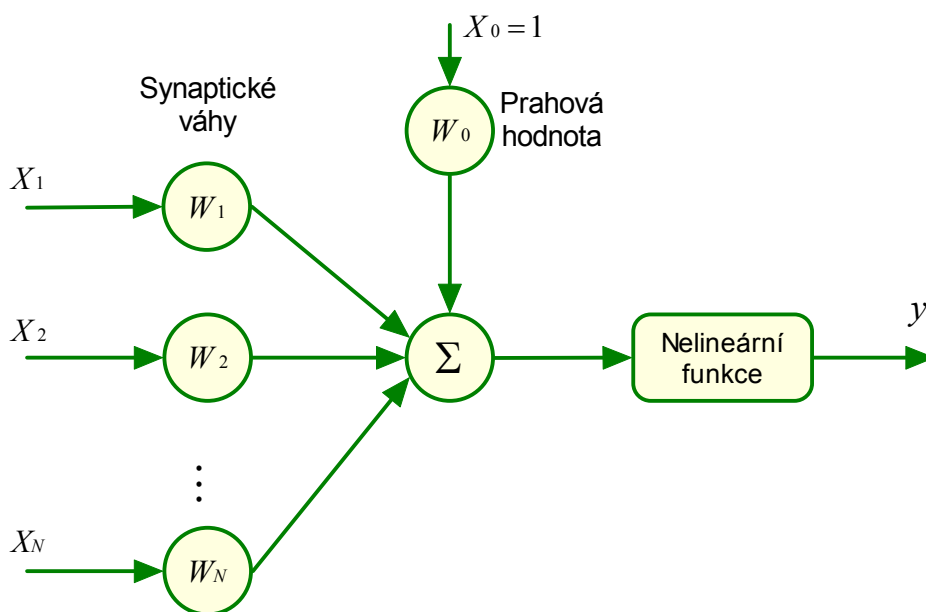
Hodnoty vah, středních hodnot a kovariančních matic jednotlivých složek modelu (5.4) se dají odhadnout například pomocí EM (Expectation Maximization) algoritmu viz. [1] odst. 5.3.3. Po dosazení těchto parametrů do GMM modelu lze vypočítat pravděpodobnosti $P(X|\lambda)$. Pro kterou třídu je výsledná pravděpodobnost větší, do té třídy bude vstupní vzor klasifikován [2].

5.4 Umělé neuronové sítě (ANN)

Umělé neuronové sítě vychází z jednoduchých modelů neuronů – funkčních buněk nervového systému živých organismů. Většina současných aplikací umělých neuronových sítí využívá selektivní a generalizační vlastnosti těchto struktur.

5.4.1 Model neuronu

První matematický model neuronu byl vytvořen McCullochem a Pittsem v roce 1943 a dodnes se používá pro běžné aplikace. Matematický model neuronu se skládá ze tří hlavních částí. Obsahuje vstupní, výstupní a funkční část. Vstupní část se skládá ze vstupů a z přiřazených nastavitelných vah (synaptické váhy). Na základě váhových koeficientů mohou být jednotlivé vstupy zvýhodňovány či potlačeny. Následující částí je výkonná jednotka, která zpracuje informace ze vstupu a vygeneruje výstupní odezvu. Třetí část je výstupní jednotka, která přivádí výstupní informace na vstup jiných neuronů. Následující obrázek popisuje funkci jednoho neuronu [8].



Obr. 5.4 Model neuronu.

Funkce jednoho neuronu je následující: Vstupní hodnoty jsou vynásobeny příslušnými váhovými koeficienty a sečtou se. Na výsledek součtu se aplikuje funkce (obecně nelineární) a výsledná hodnota funkce je přivedena na vstup jiných neuronů pomocí výstupní části. Na Obr. 5.4 je navíc vidět, že neuron má jeden zvláštní vstup, který není připojený k výstupu žádného neuronu, ale přivádí konstantní veličinu do neuronu. Tato veličina funguje jako prahová hodnota při aktivování výstupu. Když suma váženého součtu vstupů nepřesahuje prahovou hodnotu, tak se neuron neaktivuje a jeho výstup zůstane nezměněný.

5.4.2 Matematický popis neuronu

Matematicky lze funkci neuronu popsat následovně:

$$y = f\left(\sum_{i=1}^n x_i w_i + \Theta\right), \quad (5.5)$$

kde x_i je hodnota na i -tém vstupu, w_i je váha i -tého vstupu, Θ je prahová hodnota, n je celkový počet vstupů, f je obecná nelineární funkce a y je hodnota výstupu.

5.4.3 Umělé neuronové sítě

Jediný neuron není schopen vykonat příliš složitou funkci. Síla systému využívajícího umělé neurony je v síti velkého počtu neuronů. Umělá neuronová síť je vlastně pole jednoduchých výkonných prvků – neuronů. Takovéto uspořádání má velkou flexibilitu a spolehlivost. Umožňuje různě propojovat vstupy a výstupy neuronů, zvýhodnit či potlačit některé vstupy a minimalizovat vliv nesprávně fungujícího neuronu na celkový výsledek.

Neurony jsou většinou sdružovány do vrstev. Výstupy z n -té vrstvy jsou přivedeny na vstup obecně každého neuronu ve vrstvě $n+1$. První vrstva se nazývá vstupní či rozdělovací vrstva a má za úkol přijímat hodnoty z okolí pro zpracování a přivést je na vstup každého neuronu následující vrstvy. Poslední vrstva se nazývá výstupní vrstva a hodnoty na jejím výstupu jsou odezvou celého systému na vstupní vzorky. Vnitřní vrstvy se nazývají skryté vrstvy. Jejich počet závisí na složitosti funkce, kterou má síť vykonat, a na zvoleném typu sítě.

5.4.4 Typy neuronových sítí

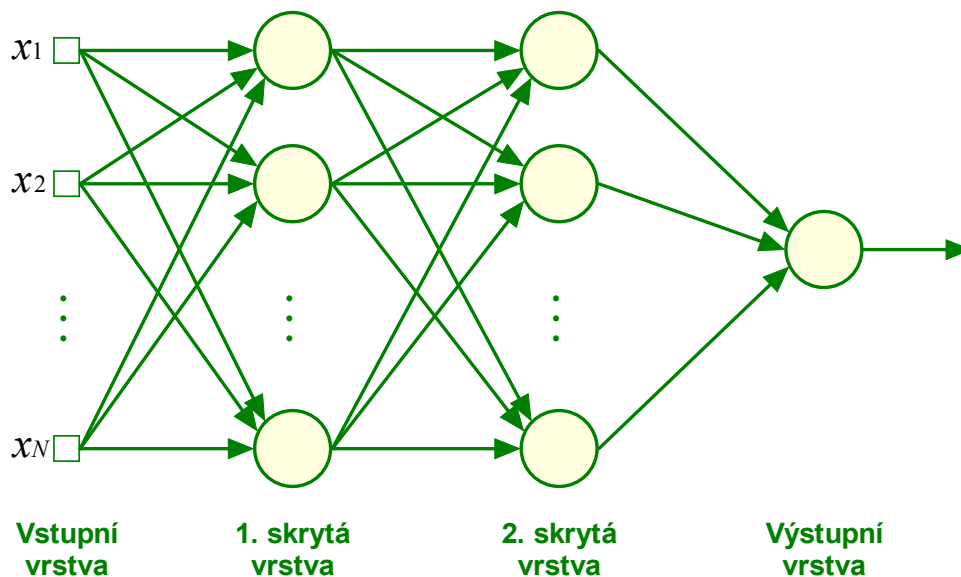
Neuronové sítě lze rozdělit do dvou hlavních skupin podle struktury: na sítě s dopředním šířením signálu a na sítě se zpětnou vazbou. V současnosti se nejčastěji používají struktury s dopředním šířením signálu, kde výstupy z jedné vrstvy jsou vedeny na vstup následující vrstvy. Výstupy z poslední, (výstupní) vrstvy jsou výstupy z celé sítě.

Neuronové sítě s dopředním šířením signálu lze rozdělit do dvou skupin podle funkce, kterou realizují, a to na lineární a nelineární. Tato funkce samozřejmě není totožná s výstupní funkcí jednoho neuronu. Sítě lineární jsou schopné realizovat lineární

matematické funkce, tj. funkce skládající se ze součtů a z násobení. Charakteristickou vlastností nelineárních neuronových sítí s dopředným šířením signálu je schopnost učení. Fáze učení předchází fázi vlastní práce a slouží k určení váhových koeficientů a k uložení informací do paměti systému. Učení se může probíhat dvěma způsoby, s učitelem a bez učitele. Při prvním způsobu je síť trénována pomocí dvojic vstupní vzorek a příslušný, očekávaný výstupní vzorek. Trénovací vstupní vzorky jsou vybrány z celkové množiny vstupních vzorků tak, aby plně popsaly všechny vlastnosti množiny důležité pro danou úlohu. V této fázi nenatréňované síti přiložíme vstupní vzorek. Na základě skutečné odezvy a očekávané odezvy se upravují váhové koeficienty. Během trénování se na vstupy sítě přivedou všechny trénovací vzorky vícekrát a navíc v náhodném pořadí. Po natréňování síť musí správně reagovat na všechny trénovací vzorky a dále má pracovat dobře i pro ostatní vzorky množiny. Aby síť pracovala dobře, potřebujeme velký počet trénovacích vzorků. Obecně platí, že čím větší je počet trénovacích vzorků, tím přesněji bude síť pracovat. Příkladem takové sítě je síť "back-propagation", která je pravděpodobně nejčastěji používaným typem [8].

5.4.5 Algoritmus učení backpropagation

Algoritmus zpětného šíření Back-propagation je jeden z nejvyužívanějších algoritmů pro trénování a učení neuronových sítí. Jedná se o vícevrstvou neuronovou síť, jejíž topologie je zobrazena na následujícím obrázku.



Obr. 5.5 Schéma neuronové sítě Back-propagation.

Jednotlivé kroky algoritmu Back-propagation [9]:

Výstupní vrstva je značena symbolem o , skrytá vrstva symbolem h , i je index výstupních neuronů a j je index vstupních neuronů. Tzn., že y_i^h je výstup i -tého neuronu skryté vrstvy a w_{ij}^0 je váha spojující i -tý neuron výstupní vrstvy s j -tým neuronem předcházející skryté vrstvy.

Krok 1. Inicializace: Všechny váhy v síti se nastaví náhodně na hodnoty v dopředném rozsahu $\langle -0,3; 0,3 \rangle$.

Krok 2. Předložení vzoru: Vybereme vzor z trénovací množiny a přiložíme na vstupy sítě. Dále se po vrstvách směrem od vstupu k výstupům sítě počítají výstupy jednotlivých neuronů dle vztahů

$$y = S\left(\sum_{i=1}^n w_i x_i + \Theta\right) \text{ a} \quad (5.6)$$

$$S(\varphi) = \frac{1}{1 + e^{-\varphi}}. \quad (5.7)$$

Krok 3. Srovnání: Nejprve se vypočte energie podle vztahu

$$E = \frac{1}{2} \sum_{i=1}^n (y_i - d_i)^2, \quad (5.8)$$

kde n je počet výstupů sítě, y_i je i -tý vstup a d_i je i -tý požadovaný výstup. Energie bude využita jako přírůstek k celkové energii počítané přes všechny vzory. Dále se počítá chyba pro výstupní vrstvu

$$\delta_i^0 = (d_i - y_i^0) y_i^0 (1 - y_i^0). \quad (5.9)$$

Krok 4. Zpětné šíření chyby: Pro všechny neurony ve vrstvě se vypočte

$$w_{ij}^l(t) = \eta \delta_i^l(t) y_j^{l-1}(t) + \alpha \Delta w_{ij}^l(t-1), \quad (5.10)$$

$$\Theta_i^l(t) = \eta \delta_i^l(t) + \alpha \Delta \Theta_i^l(t-1), \quad (5.11)$$

kde α znamená setrvačnost učení a η je rychlost učení či strmost kroku. Podle vztahu

$$\delta_i^{h-1} = y_i^{h-1} (1 - y_i^{h-1}) \sum_{k=1}^n w_{ki}^h \delta_k^h \quad (5.12)$$

se zpětně šíří chyba do vrstvy, která je blíže vstupům. Na závěr tohoto kroku se modifikují váhy.

$$w_{ij}^l(t+1) = w_{ij}^l(t) + \Delta w_{ij}^l(t), \quad (5.13)$$

$$\Theta_i^l(t+1) = \Theta_i^l(t) + \Delta \Theta_i^l(t). \quad (5.14)$$

Tento krok se opakuje pro všechny vrstvy sítě tak, že se začíná vrstvou výstupní ($l=h$), pak následují skryté vrstvy ($l=h$). V případě, že se zpracovává skrytá vrstva, která je nejbližší vstupní vrstvě, nahradí se y_i^{h-1} v rovnici pro výpočet Δw za odpovídající vstupní hodnotu, tedy za x_j .

Krok 5. Ukončení výběru vzorů z trénovací množiny: Jestliže se předložily síti všechny vzory z trénovací množiny, pokračuje se krokem 6. Jinak se postup opakuje od kroku 2.

Krok 6. Ukončení procesu učení: Jestliže energie neuronové sítě za poslední epochu byla menší než předem zvolené kritérium, pak se učení ukončí. Jinak se pokračuje krokem 2 [9].

6 Detektor logopedických vad v řeči

Metody pro detekci logopedických vad v řeči byly vyvíjeny v rámci grantového projektu ve spolupráci s Ing. Hichamem Atassim z ústavu telekomunikací FEKT VUT a se studentem magisterského programu MTIT FEKT VUT Bc. Lubomírem Dušilem.

Navržený detektor logopedických vad se skládá ze dvou hlavních bloků, a to z výběru hlásky ze slova a z klasifikace vybrané hlásky do třídy vadné nebo správné výslovnosti. Princip detektoru naznačuje **Chyba! Nenalezen zdroj odkazů.** Jednotlivé části detektoru byly vytvořeny v programovém prostředí Matlab. Jelikož je detekce logopedických vad velice rozsáhlá problematika a pro klasifikaci jedné hlásky je třeba získat velké množství nahrávek obsahujících konkrétní hlásku, byl navržený detektor koncipován pouze pro hlásky [r], [ř], [s] a [š]. V těchto hláskách se logopedické vady vyskytují nejčastěji.



Obr. 6.1 Princip navrženého detektoru.

6.1 Předzpracování

V bloku předzpracování je nejprve nahrávka digitalizována a po té je z nahrávky vybrána ta část, která obsahuje řeč.

Digitalizace signálu je provedena již v konkrétním rekordéru, proto není nutné se zde digitalizací zabývat.

Pro nahrávání a předzpracování byl vytvořen skript nahraj.m, který nahraje 5 vteřin akustického signálu z mikrofону se vzorkovací frekvencí 16 kHz a na základě energetické analýzy signálu vybere pouze tu část signálu, která obsahuje řeč.

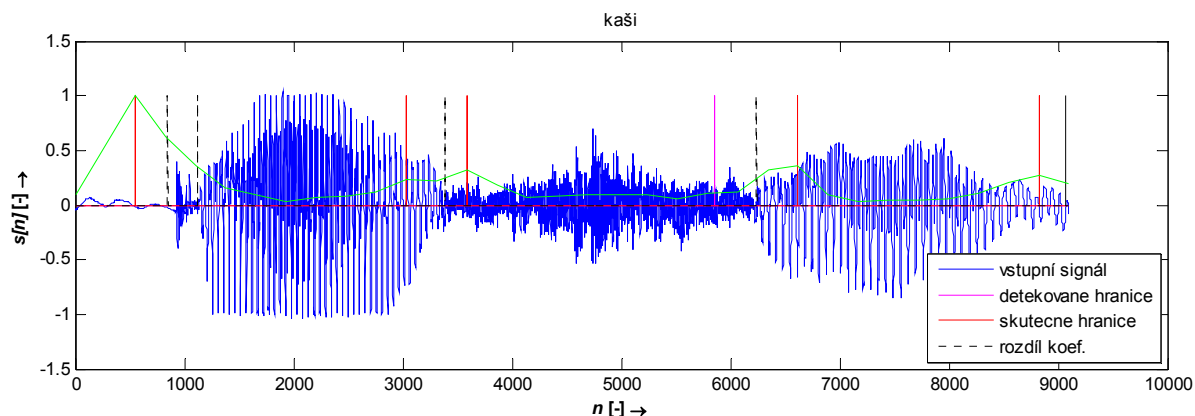
6.2 Segmentace slova a výběr hlásky

Nejprve musí být stanoveno, která hláska z nahraného slova má být klasifikována. Výběr hlásky je možné rozdělit na dva kroky. Nejprve se nahrané slovo rozdělí na jednotlivé hlásky (segmentace) metodou sledování rozdílnosti příznaků, která je popsána výše v kapitole 4. Po rozdělení slova na jednotlivé hlásky je třeba vybrat hlásku, která má být klasifikována. Jelikož výsledná křivka rozdílnosti může obsahovat více maxim, než je skutečný počet hranic mezi hláskami (nadsegmentování), nebo méně maxim (podsegmentování), není úkol výběru hlásky jednoduchý. Pro tento úkol byl použit algoritmus GMM, který je podrobněji popsán výše v kapitole **Chyba! Nenalezen zdroj odkazů.** Algoritmus GMM na základě statistického rozpoznání vzoru vrátí pravděpodobnosti výskytu natrénované hlásky v signálu. GMM je natrénován MFCC příznaky, které jsou vypočítány z trénovacích nahrávek hlásek [r], [ř], [s], [š]. Ze vstupního signálu jsou vypočítány MFCC příznaky a je vypočítána křivka podobnosti jednotlivých rámců trénovacích dat s rámci vstupního signálu. Podle toho, kde se vyskytuje maximální pravděpodobnost výskytu, je vybrán segment slova. Vybraný segment slova odpovídá nejpravděpodobnější poloze vybrané hlásky ve slově. Výběr trénovacích nahrávek hlásky má velký vliv na rozpoznání hlásky v signálu. Jelikož není známo, zda vstupní slovo obsahuje hlásku s vadnou nebo se správnou výslovností, musí být natrénovány oba typy hlásky.

Pro výběr hlásky z nahraného slova byl vytvořen skript `vyber_hlasku.m`. Vstupem této funkce je slovo a výstupem je hláska. Parametry, které byly nastaveny, jsou následující (vzorkovací frekvence vstupního signálu je 16 kHz):

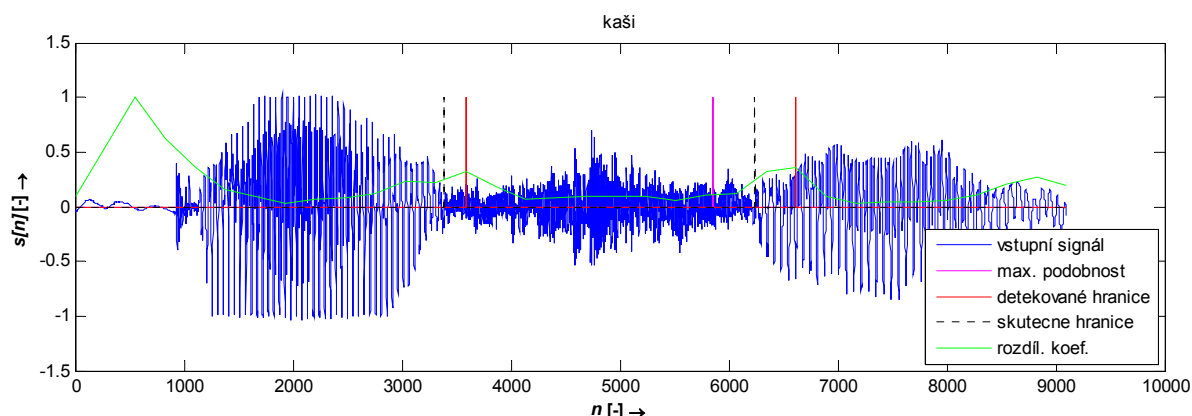
- hláska, která má být na výstupu – vstupní parametr,
- parametry dělení na rámce: délka rámce – 512 (32 ms),
překryv rámců – 256 (16 ms),
- parametry MFCC pro GMM: délka rámce – 256 (16 ms),
překryv rámců – 128 (8 ms), počet příznaků – 24,
- hodnota prahu pro prahování křivky rozdílnosti – 0,2.

Příklad rozdělení slova na segmenty ukazuje Obr. 6.2, kde je modrou barvou zobrazen vstupní signál, zelenou barvou křivka rozdílnosti a červenou barvou detekované hranice hlásek odpovídající maximum křivky rozdílnosti, které převyšují stanovený práh. Skutečné hranice hlásek ve slově jsou vyznačeny přerušovaně, černou barvou.



Obr. 6.2 Příklad rozdělení slova na segmenty.

Příklad výběru hlásky ze segmentovaného slova naznačuje Obr. 6.3, kde je opět segmentovaný signál značen modrou barvou, křivka rozdílnosti zelenou barvou. Maximální podobnost signálu s hledanou hláskou je vyznačena fialovou barvou a detekované hranice hlásky barvou červenou. Skutečné hranice hledané hlásky jsou značeny přerušovaně černou barvou.

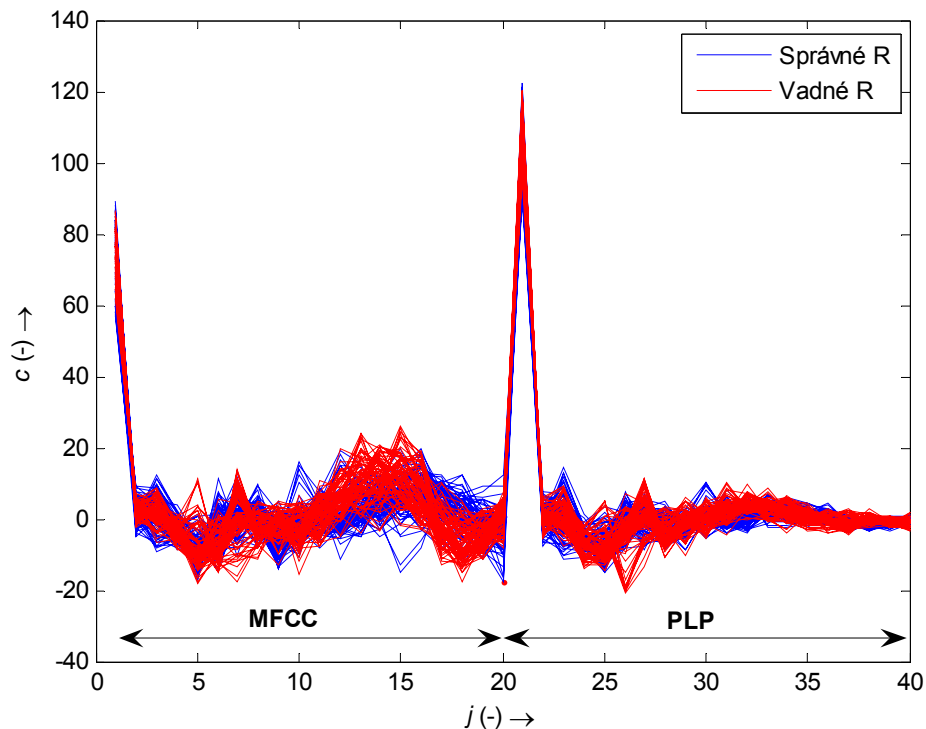


Obr. 6.3 Příklad výběru hranic hlásky

6.3 Klasifikace

Pro klasifikaci hlásky do tříd správné nebo vadné výslovnosti je možné aplikovat různé metody rozpoznání vzoru. V této práci bude na základě testování vybrána jedna metoda z metod KNN, SVM, ANN a GMM, které patří mezi nejpoužívanější metody rozpoznání vzoru. Tyto metody jsou popsány v kapitole 5.

Hlásky, která má být klasifikována je segmentována na rámce. Pro každý rámec jsou vypočítány příznaky hlasového traktu. Tyto příznaky jsou vypočítány metodami MFCC, PLP, nebo jejich první a druhou derivací, případně jejich kombinacemi. Příklad kombinace MFCC a PLP příznaků pro správnou hlásku [r] (modrá) a vadnou hlásku [r] (červená) ukazuje následující obrázek. Výstupy z MFCC a PLP jsou zařazeny za sebou (MFCC:1-20, PLP:21-40). Aby byl obrázek přehlednější, byl omezen počet zobrazených segmentů.



Obr. 6.4 MFCC a PLP kepstrum pro správnou a vadnou hlásku [r].

Extrahované příznaky jedné hlásky tvoří matici o rozměru $M \times N$, kde M značí počet rámců, na které byla hláska rozdělena a N je počet příznaků v jednom rámcí. Jelikož má každá nahrávka hlásky jinou délku, je i počet rámců v matici příznaků různý. Proto je před samotnou klasifikací zařazena některá z metod pro výběr příznaků. Pro výběr příznaků mohou být použity různé postupy, v rámci testování bude vybrána jedna z následujících metod.

- Každý rámeček je klasifikován zvlášť.

V tomto případě je pro každý rámeček určena třída vadné nebo správné výslovnosti. Podle toho, která třída převažuje, je určena třída klasifikované hlásky. Pokud je do obou tříd klasifikován stejný počet rámců, je vybrána třída vadné výslovnosti. Je totiž vhodnější, aby byla správná hláska klasifikována jako špatná, než aby byla špatná hláska klasifikována jako správná.

- Rámce jsou řazeny pod sebe.

Jednotlivé rámce s příznaky jsou zařazeny pod sebe, čímž vznikne jeden vektor příznaků pro jednu hlásku. Vzniklé vektory příznaků jednotlivých hlásek jsou dále upraveny na stejnou délku doplněním nulami a tyto vektory jsou klasifikovány.

- Rámce jsou průměrovány.

Z každého příznaku všech rámců je vypočítán aritmetický průměr. Vznikne tak jeden rámec obsahující průměrné hodnoty příznaků všech rámců, které charakterizují danou hlásku. Výsledný rámec je klasifikován.

- Mezi další možné metody výběru příznaků patří například metoda PCA nebo LDA.

Vybrané příznaky jsou dále klasifikovány použitím jednoho z klasifikátorů. Na výstupu klasifikátoru je matice výsledků. Tato matice se podobá Target matici, která určuje třídu trénovacích dat (viz. následující odstavec). Tvar matice výsledků je dán podle toho, jaká byla zvolena metoda pro výběr příznaků. Matice výsledků určuje, do které třídy byl testovaný vzor klasifikován.

6.3.1 Trénování klasifikátoru a tvorba testovací a trénovací databáze.

Každý klasifikátor musí být před vlastní klasifikací vzoru natrénován trénovacími vzory. Výjimkou je klasifikátor k -NN, kde trénování a klasifikace probíhají současně. Trénovací příznaky jsou vypočítány stejným způsobem jako příznaky pro testování. Pro úsporu paměti pro trénovací data a pro snížení výpočetní náročnosti je vhodné místo trénovacích nahrávek uložit extrahované příznaky. Poté se již při trénování klasifikátoru nemusí příznaky počítat, jelikož jsou načteny přímo z paměti.

Důležitou podmínkou pro správné fungování klasifikátoru je vytvoření kvalitní databáze pro trénování klasifikátoru. V rámci této práce byla vytvořena databáze nahrávek hlásek [r], [ř], [s], [š] se správnou a s vadnou výslovností. Hlázky s vadnou výslovností byly „nastříhány“ z nahrávek dětí z logopedické poradny. Hlázky se správnou výslovností měly být původně „vystříhány“ z nahrávek dětí s dobrou artikulací, například dětí z pěveckého sboru. Bohužel se tyto nahrávky nepodařilo získat. Proto byly použity nahrávky z dostupných zdrojů. Část nahrávek správné výslovnosti byla „vystříhána“ z nahrávek dětí z logopedické poradny a to za předpokladu, že pokud dítě nemá vadu některé hlásky, vyslovuje ji správně. Počet těchto nahrávek byl nedostatečný, proto byl doplněn nahrávkami dětí z mateřské a základní školy. Přehled počtu nahrávek v databázi zobrazuje tabulka 6.1. Část nahrávek z databáze je použita pro trénování a část pro testování klasifikátorů. Klasifikátor je natrénován na tu hlásku z vytvořené databáze, kterou má klasifikovat. Pro databázi nahrávek je vytvořena tzv. Target matice, která „říká“ klasifikátoru, která nahrávka spadá do třídy 1 (hláska bez vady) a která do třídy 0 (hláska s vadou).

Hláška	Vadná výslovnost		Správná výslovnost	
	Počet mluvčích	Počet nahrávek	Počet mluvčích	Počet nahrávek
R	37	106	46	144
Ř	40	133	25	84
S	52	225	34	196
Š	41	124	41	122

Tab. 6.1 Přehled počtu nahrávek v databázi

6.4 Učíme se správně mluvit

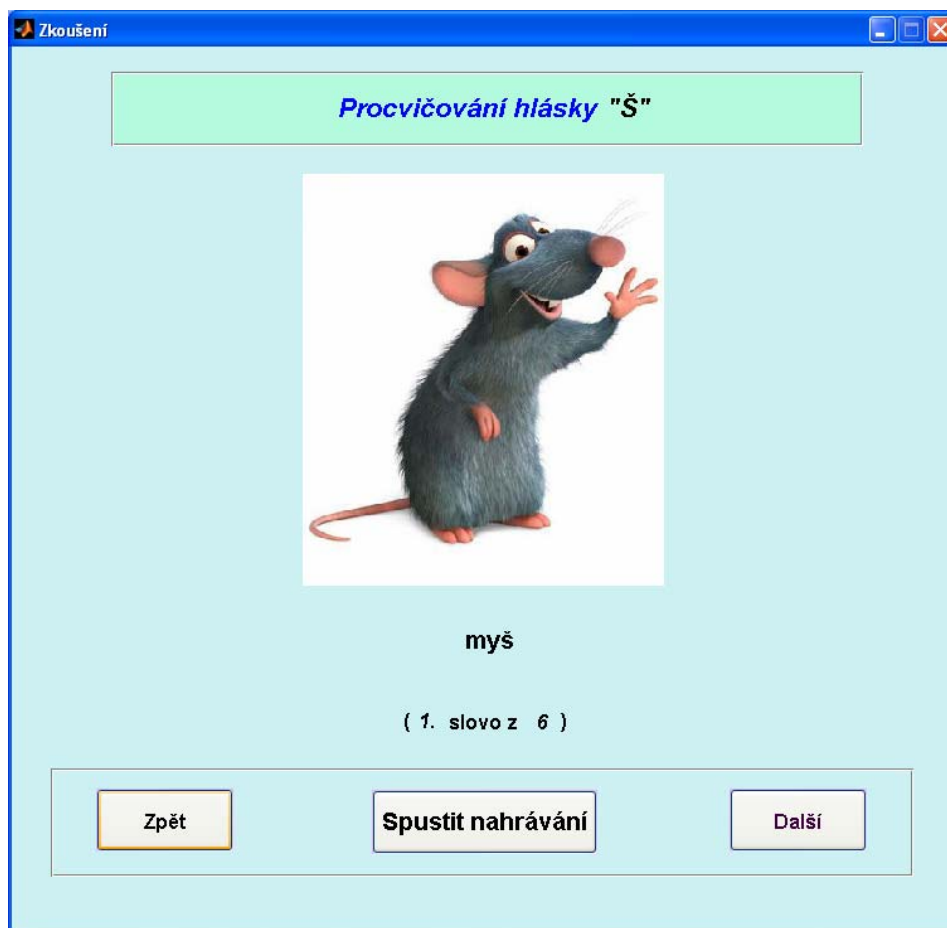
Jako návrh grafické podoby programu bylo vytvořeno uživatelské prostředí „Učíme se správně mluvit“.

Uživatelské prostředí detektoru logopedických vad v řeči bylo vytvořeno v grafickém prostředí GUIDE programu Matlab. Průvodce tvorbou GUI (GUIDE) programu Matlab je grafické interaktivní vývojové prostředí obsahující všechny grafické objekty, kterými je možno ovládat běh aplikace. Generuje automaticky základní zdrojový kód pro ovládání [13].



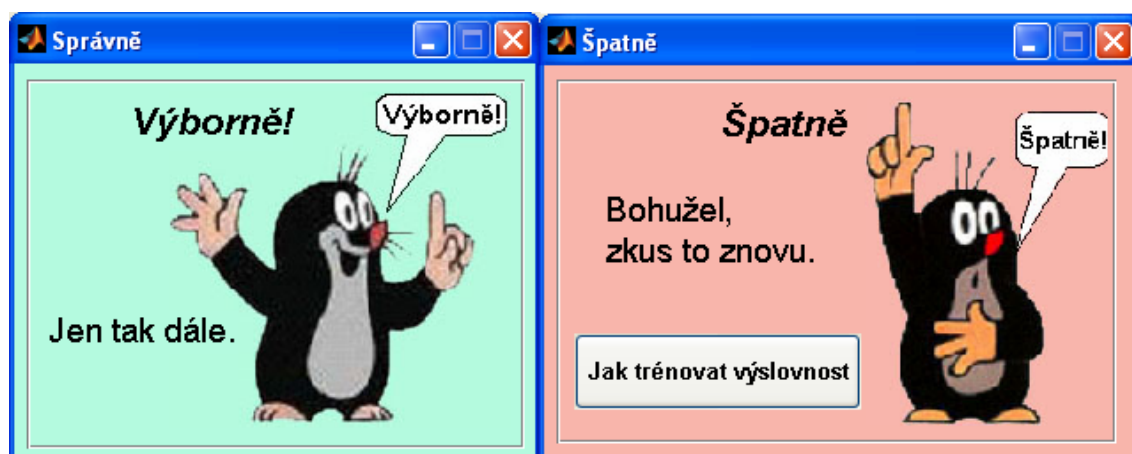
Obr. 6.5 Hlavní nabídka programu

Hlavní nabídku programu zobrazuje Obr. 6.5. V této nabídce je možné zobrazit informace o programu a stručný návod k použití. Dále zde uživatel zvolí hlásku, která má být procvičována. Tlačítkem „Začít procvičovat“ se spustí procvičování vybrané hlásky. Program se ukončí tlačítkem „Ukončit program“.



Obr. 6.6 Procvičování hlásky [š].

Po stisknutí tlačítka „Začít procvičovat“ se otevře okno „Zkoušení“, viz. Obr. 6.6. **Chyba! Nenalezen zdroj odkazů.** Zde je zobrazen obrázek pro slovo obsahující hlásku, která byla zvolena v hlavní nabídce programu. Stiskem tlačítka „Spustit nahrávání“ se nahraje 5 vteřin signálu z mikrofону. Pokud je nahrávka slabá, je uživatel vyzván k opakování nahrávání. Po ukončení nahrávání je výslovnost zvolené hlásky vyhodnocena. Způsob vyhodnocení výslovnosti hlásky je zobrazen v Obr. 6.7. Jelikož se většinou logopedické vady rozpoznávají u dětí v předškolním věku, které ještě neumějí číst, bylo by vhodnější jako vyhodnocení výslovnosti použít animaci se zvukovým doprovodem. Dítě by tak ihned pochopilo, zda byla hláska vyslovena dobře nebo špatně, a bylo by motivováno k lepší výslovnosti. V případě vyhodnocení hlásky, jako špatně vyslovená hláska je možné stisknout tlačítko „Jak trénovat výslovnost“. Tím se zobrazí text, který stručně popisuje základní cviky pro zdokonalení artikulace. Krok nahrávání lze pro totéž slovo neomezeně opakovat nebo je možné tlačítkem „Další“ přejít na jiný obrázek, který opět vystihuje slovo obsahující hlásku zvolenou v hlavní nabídce. Tlačítkem „Zpět“ se lze vrátit na předchozí slovo.



Obr. 6.7 Vyhodnocení výslovnosti hlásky

7 Testování algoritmu pro výběr hlásky ze slova

Pro testování automatické segmentace byl vytvořen skript `uspesnost_segmentace.m`, který vrací statistiku úspěšnosti segmentace slov a nalezení hlásky ve slově. Aby bylo možné testovat úspěšnost automatické segmentace, musela být vytvořena databáze nahrávek slov, které byly ručně značkovány (labeling). Značky hranic hlásek jsou pro každé testované slovo uloženy v textovém souboru. Vytvořený skript v jednom cyklu načte nahrávku slova a skutečné hranice hlásek z textového souboru. Slovo je segmentováno na hlásky (viz. kapitola 4) a detekované hranice hlásek jsou porovnávány se skutečnými. Pro vyhodnocení úspěšnosti segmentace jsou vypočítány následující hodnoty.

- Počet správně detekovaných hranic v procentech P_c .

Za správně detekovanou hranici hlásky se považuje hranice, která se od skutečné hranice liší maximálně o 30 ms. Procentuální počet správně detekovaných hranic P_c definuje vzorec

$$P_c = 100 \cdot \frac{S_c}{S_t} , \quad (7.1)$$

kde S_c je počet správně detekovaných hranic ve slově a S_t je počet skutečných hranic hlásek ve slově.

- Počet chybně detekovaných hranic v procentech P_m .

Za chybně detekovanou hranici hlásky se považuje hranice, která se od skutečné hranice liší o více než 30 ms. Procentuální počet chybně detekovaných hranic P_m definuje vzorec

$$P_m = 100 - P_c = 100 \cdot \left(1 - \frac{S_c}{S_t} \right) . \quad (7.2)$$

- Nadsegmentování / podsegmentování.

Jako nadsegmentování se označuje případ, kdy je ve slově detekováno více hranic hlásek než je ve skutečnosti. Naopak podsegmentování vyjadřuje případ, kdy je ve slově detekováno méně hranic hlásek. Nadsegmentování nebo podsegmentování vyjadřuje hodnota S_{ins} a její procentuální vyjádření D . Pokud je hodnota kladná, jedná se o nadsegmentování. Záporná hodnota vyjadřuje podsegmentování. Hodnoty S_{ins} a D jsou vypočítány podle vzorců

$$S_{ins} = S_d - S_t , \quad (7.3)$$

$$D = 100 \cdot \left(\frac{S_d}{S_t} - 1 \right). \quad (7.4)$$

Hodnoty P_m , P_c , S_{ins} a D jsou vypočítány pro všechny testovací slova a ze všech hodnot je vypočítán aritmetický průměr [11].

Pro testování segmentace slov bylo v programovém prostředí Elan značkováno (labelováno) 36 nahrávek slov obsahujících hlásku [r], [ř], [s], [š]. Výsledky testování úspěšnosti segmentace jsou uvedeny v Tab. 7.1.

Slovo	P_c [%]	P_m [%]	S_{ins} [-]	D [%]
kaši	80,0	20,0	0	0,0
kaši2	80,0	20,0	0	0,0
kos	50,0	50,0	2	50,0
křeček	80,0	20,0	1	20,0
kuřátko	87,5	12,5	8	100,0
kuře	71,4	28,6	6	85,7
les	100,0	0,0	15	375,0
les2	100,0	0,0	10	250,0
liška	100,0	0,0	6	100,0
mrkev	83,3	16,7	3	50,0
pase	80,0	20,0	-1	-20,0
pejsek	71,4	28,6	-2	-28,6
pejsek1	57,1	42,9	-2	-28,6
pes	100,0	0,0	0	0,0
pračka	42,9	57,1	0	0,0
rajčata	75,0	25,0	6	75,0
ramínko	77,8	22,2	6	66,7
rampouchy	62,5	37,5	3	37,5
ruce	100,0	0,0	5	100,0
ručník	66,7	33,3	3	33,3
rukavice2	85,7	14,3	7	100,0
růže	33,3	66,7	-1	-16,7
rybář	100,0	0,0	7	140,0
rýč	25,0	75,0	7	175,0
ředkvičky	100,0	0,0	10	166,7
řeka	55,6	44,4	2	22,2
řepa	66,7	33,3	2	22,2
řetaz	80,0	20,0	8	160,0
řeže1	85,7	14,3	10	142,9
sloni	60,0	40,0	-1	-20,0
sněhulák	0,0	100,0	0	0,0
sněhulák2	60,0	40,0	3	30,0
sova1	80,0	20,0	5	100,0
střecha	20,0	80,0	-1	-20,0
sval	83,3	16,7	10	166,7
vaří	80,0	20,0	0	0,0
Průměr	71,7	28,3	3,8	67,6

Tab. 7.1 Výsledky testování automatické segmentace pro 36 slov.

Z tabulky výsledků vychází průměrná úspěšnost segmentace 71,7 %, s nadvzorkováním 67,6 %, což odpovídá průměrně o 3,8 nalezených hranic ve slově více než ve skutečnosti.

Pro určení úspěšnosti detekce hranic hlásky jsou vypočítány následující hodnoty:

- Odchylka nalezených hranic hlásky od správných E .

Odchylka E udává, o kolik milisekund jsou vzdáleny nalezené hranice hlásky od skutečných. Výpočet odchylky definuje vzorec

$$E = \frac{|h_s - h_d|}{F_{vz}} \cdot 1000, \quad (7.5)$$

kde h_s je poloha skutečné hranice hlásky ve vzorcích, h_d je poloha detekované hranice hlásky ve vzorcích a F_{vz} je vzorkovací frekvence signálu. Pokud je odchylka obou hranic hlásky menší než 50 ms, je detekce hranic považována za úspěšnou.

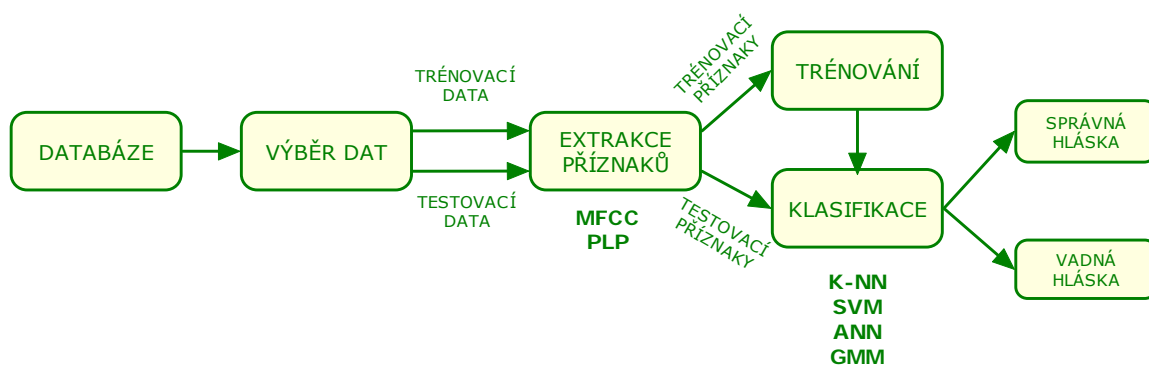
Hlásky [r] a [ř] byly touto metodou nalezeny s úspěšností pod 50 %, hlásky [s] a [š] s úspěšností kolem 90 %. Na tyto výsledky má značný vliv správné natrénování GMM klasifikátoru. Trénovací nahrávky hlásek [r] a [ř] obsahovaly mnoho přechodových dějů, proto byla úspěšnost jejich nalezení velice nízká.

Pro použití uvedeného algoritmu pro výběr hlásky z rozsegmentovaného slova je vhodnější, když je slovo nadsegmentováno než podsegmentováno. Pokud by bylo slovo podsegmentováno, mohla by být místo jedné hlásky označena větší část signálu.

8 Testování klasifikátorů

Pro testování klasifikátorů byla upravena již existující aplikace, která byla původně vytvořena na ústavu telekomunikací pro rozpoznání emocí mluvčího.

Proces testování klasifikátoru naznačuje blokové schéma v **Chyba! Nenalezen zdroj odkazů.**



Obr. 8.1 Blokové schéma testování klasifikátoru.

8.1 Výběr dat

Nejprve je načtena databáze nahrávek hlásky, která má být klasifikována. Z této databáze jsou vybrány hlásky na trénování a hlásky na testování klasifikátoru. Tento výběr může být proveden různými způsoby. Jelikož je databáze trénovacích nahrávek vytvořena od velkého počtu mluvčích, jsou klasifikátory nezávislé na mluvčím. Použitá aplikace umožňuje nastavit poměr počtu vybraných hlásek pro trénování ku počtu vybraných hlásek pro testování.

8.2 Extrakce příznaků

V aplikaci je možné zvolit, kterými metodami budou vypočítány příznaky z načtených nahrávek.

Během testování jsou použity kombinace příznaků metod podle následujícího přehledu.

- MFCC
- MFCC, Δ MFCC
- MFCC, Δ MFCC, $\Delta\Delta$ MFCC
- PLP
- PLP, Δ PLP
- PLP, Δ PLP, $\Delta\Delta$ PLP
- MFCC, PLP

- MFCC, Δ MFCC, PLP, Δ PLP
- MFCC, Δ MFCC, $\Delta\Delta$ MFCC, PLP, Δ PLP, $\Delta\Delta$ PLP

Pro nastavení parametrů metod pro extrakci příznaků bylo vypočteno několik předběžných testů a podle výsledků úspěšnosti klasifikace pro různé parametry byly nastaveny následující hodnoty, (vzorkovací frekvence vstupního signálu je 16 kHz):

- délka rámce: 256 (16 ms),
- překryv rámců: 128 (8 ms),
- počet extrahovaných příznaků: 20,
- minimální frekvence filtrů: 0 Hz,
- maximální frekvence filtrů: 4000 Hz.

Dále aplikace umožňuje nastavení metody pro výběr příznaků (viz. kapitola 6.3.) Podle předběžných testů vycházela nejlépe metoda kdy je každý rámeček klasifikován zvlášť. Proto byly další testy prováděny pomocí této metody.

8.3 Klasifikace

Testovací aplikace umožňuje volbu jedné z metod pro výběr příznaků, jejichž přehled je uveden v předchozím odstavci. Dále je možné vybrat jeden z klasifikátorů a nastavit jeho parametry.

8.3.1 Přehled nastavení parametrů jednotlivých klasifikátorů:

***k*-NN:**

- počet nejbližších sousedů k ,

SVM:

- jádrová funkce,

ANN:

- počet vrstev,
- počet neuronů ve vrstvách,

GMM:

- počet Gaussových funkcí.

Aplikace umožňuje nastavit, kolik opakování má proběhnout pro testování jednoho klasifikátoru. Během jednoho opakování proběhne klasifikace všech vzorů, které byly vybrány pro testování. V následujícím opakování je proveden nový výběr dat. To znamená, že jsou vybrány jiné nahrávky pro testování a trénování a že znovu proběhne klasifikace všech testovacích vzorů. Pro každé opakování je vypočítána matice úspěšnosti klasifikace. Tato matice vyjadřuje procentuální počet správně zařazených a procentuální počet chybně zařazených vzorů. Z celkového počtu opakování je

následně vypočítána matice, která vyjadřuje průměrnou procentuální úspěšnost klasifikace vzoru pro všechny opakování.

8.4 Výsledky testování klasifikátoru k -NN

Klasifikátor k -NN byl testován pro $k = 1, 3, 5, 19$. Hodnota k musí být lichá. Pokud by byla sudá, mohl by klasifikátor dojít ke stejnému počtu nejbližších sousedů patřících do jedné i druhé třídy. V následujících tabulkách jsou zobrazeny výsledky testování klasifikátoru k -NN vždy pro jednu hlásku a jednu kombinaci příznaků z přehledu, který byl uveden výše.

Testovaná hláska: [r]

Příznaky	MFCC			
k	1	3	5	19
Bez vady [%]	89	84,6	84,6	84,2
Vada [%]	69,4	68	64,7	62,8
Průměr [%]	79,2	76,3	74,65	73,5
Příznaky	MFCC, Δ MFCC			
k	1	3	5	19
Bez vady [%]	75,2	77,2	74,4	78
Vada [%]	65,6	70,4	66,8	78
Průměr [%]	70,4	73,8	70,6	78
Příznaky	MFCC, Δ MFCC, $\Delta\Delta$ MFCC			
k	1	3	5	19
Bez vady [%]	79	73,2	75,2	74,2
Vada [%]	69,6	74,2	75	74
Průměr [%]	74,3	73,7	75,1	74,1
Příznaky	PLP			
k	1	3	5	19
Bez vady [%]	74	74,2	68,4	70,4
Vada [%]	65,6	67,6	65,6	62
Průměr [%]	69,8	70,9	67	66,2
Příznaky	PLP, Δ PLP			
k	1	3	5	19
Bez vady [%]	75,2	74	68,6	67,75
Vada [%]	71,4	68,4	72,4	72,5
Průměr [%]	73,3	71,2	70,5	70,1
Příznaky	PLP, Δ PLP, $\Delta\Delta$ PLP			
k	1	3	5	19
Bez vady [%]	83,25	80	76,4	75,6
Vada [%]	61,25	65,6	70	73,4
Průměr [%]	72,25	72,8	73,2	74,5
Příznaky	MFCC, PLP			
k	1	3	5	19
Bez vady [%]	79	71	80,2	73,4
Vada [%]	60	69	74,4	56,8
Průměr [%]	69,5	70	77,3	65,1

Příznaky	MFCC, Δ MFCC, PLP, Δ PLP			
k	1	3	5	19
Bez vady [%]	75,6	78	79	69
Vada [%]	63,4	73,4	68,8	62
Průměr [%]	69,5	75,7	73,9	65,5
Příznaky	MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP			
k	1	3	5	19
Bez vady [%]	77,8	78,8	80	78,6
Vada [%]	66,6	61,2	72,2	64,6
Průměr [%]	72,2	70	76,1	71,6

Tab. 8.1 Výsledky testů k -NN pro hlásku [r].

Testovaná hláska: [ř]

Příznaky	MFCC			
k	1	3	5	19
Bez vady [%]	85,8	88,3	87,6	83,9
Vada [%]	79,5	79	78,3	79,7
Průměr [%]	82,65	83,65	82,95	81,8
Příznaky	MFCC, Δ MFCC			
k	1	3	5	19
Bez vady [%]	85,9	83,8	85,1	83,2
Vada [%]	73,2	72,5	68,8	65,8
Průměr [%]	79,55	78,15	76,95	74,5
Příznaky	MFCC, Δ MFCC, Δ Δ MFCC			
k	1	3	5	19
Bez vady [%]	90,8	90,3	87,7	78,9
Vada [%]	69,5	67	69	68,2
Průměr [%]	80,15	78,65	78,35	73,55
Příznaky	PLP			
k	1	3	5	19
Bez vady [%]	83,8	88,9	89	87,7
Vada [%]	86,5	82,5	84,5	84,6
Průměr [%]	85,15	85,7	86,75	86,15
Příznaky	PLP, Δ PLP			
k	1	3	5	19
Bez vady [%]	85,1	89,5	82	79,5
Vada [%]	85,2	83,3	81,9	82,5
Průměr [%]	85,15	86,4	81,95	81
Příznaky	PLP, Δ PLP, Δ Δ PLP			
k	1	3	5	19
Bez vady [%]	84,6	90,2	88,4	82
Vada [%]	84	83,8	84	82,5
Průměr [%]	84,3	87	86,2	82,25

Příznaky	MFCC, PLP			
k	1	3	5	19
Bez vady [%]	83,3	88,9	94	87
Vada [%]	82,1	85,2	79,5	86,3
Průměr [%]	82,7	87,05	86,75	86,65
Příznaky	MFCC, Δ MFCC, PLP, Δ PLP			
k	1	3	5	19
Bez vady [%]	87,1	87,1	85,1	82
Vada [%]	83,2	81,3	80,1	82
Průměr [%]	85,15	84,2	82,6	82
Příznaky	MFCC, Δ MFCC, Δ Δ MFCC, PLP, Δ PLP, Δ Δ PLP			
k	1	3	5	19
Bez vady [%]	85,7	91,4	79,5	84,3
Vada [%]	82,1	83,9	88,2	84
Průměr [%]	83,9	87,65	83,85	84,15

Tab. 8.2 Výsledky testů k -NN pro hlásku [ř].

Testovaná hláska: [s]

Příznaky	MFCC			
k	1	3	5	19
Bez vady [%]	82,3	82,5	82,8	84,3
Vada [%]	74,9	76,7	74,5	80,8
Průměr [%]	78,6	79,6	78,65	82,55
Příznaky	MFCC, Δ MFCC			
k	1	3	5	19
Bez vady [%]	80,3	79,9	79	77,7
Vada [%]	75,4	76,6	77,8	78,8
Průměr [%]	77,85	78,25	78,4	78,25
Příznaky	MFCC, Δ MFCC, Δ Δ MFCC			
k	1	3	5	19
Bez vady [%]	69,5	69,2	69,5	67,4
Vada [%]	76,6	77,4	75,8	82,4
Průměr [%]	73,05	73,3	72,65	74,9
Příznaky	PLP			
k	1	3	5	19
Bez vady [%]	82,1	87,2	86,2	91,8
Vada [%]	67,5	72,6	76,8	72,8
Průměr [%]	74,8	79,9	81,5	82,3
Příznaky	PLP, Δ PLP			
k	1	3	5	19
Bez vady [%]	83,4	84,8	82,4	87,4
Vada [%]	65,4	65	71,2	69,8
Průměr [%]	74,4	74,9	76,8	78,6

Příznaky	PLP, Δ PLP, $\Delta\Delta$ PLP			
k	1	3	5	19
Bez vady [%]	82,8	83	81,4	87,2
Vada [%]	66,8	71,6	72	65,2
Průměr [%]	74,8	77,3	76,7	76,2
Příznaky	MFCC, PLP			
k	1	3	5	19
Bez vady [%]	80,8	83,6	82,6	85,8
Vada [%]	64,2	70,4	70,8	72
Průměr [%]	72,5	77	76,7	78,9
Příznaky	MFCC, Δ MFCC, PLP, Δ PLP			
k	1	3	5	19
Bez vady [%]	86,8	89,6	85,8	91,4
Vada [%]	64,6	66,2	67,8	67,2
Průměr [%]	75,7	77,9	76,8	79,3
Příznaky	MFCC, Δ MFCC, $\Delta\Delta$ MFCC, PLP, Δ PLP, $\Delta\Delta$ PLP			
k	1	3	5	19
Bez vady [%]	85,4	87,2	88,8	87
Vada [%]	68,6	71,2	66	65
Průměr [%]	77	79,2	77,4	76

Tab. 8.3 Výsledky testů k -NN pro hlásku [s].

Testovaná hláska: [š]

Příznaky	MFCC			
k	1	3	5	19
Bez vady [%]	79,7	87,3	86,3	85,4
Vada [%]	66,3	62,6	61,9	60,6
Průměr [%]	73	74,95	74,1	73
Příznaky	MFCC, Δ MFCC			
k	1	3	5	19
Bez vady [%]	85,9	88,6	90,1	87,3
Vada [%]	56,3	65,1	62,6	52,7
Průměr [%]	71,1	76,85	76,35	70
Příznaky	MFCC, Δ MFCC, $\Delta\Delta$ MFCC			
k	1	3	5	19
Bez vady [%]	81	90,3	86,4	91,5
Vada [%]	62,5	50,4	58,4	52,2
Průměr [%]	71,75	70,35	72,4	71,85
Příznaky	PLP			
k	1	3	5	19
Bez vady [%]	88,5	84,2	86,1	91,1
Vada [%]	63	63,3	63,3	65,6
Průměr [%]	75,75	73,75	74,7	78,35

Příznaky	PLP, Δ PLP			
k	1	3	5	19
Bez vady [%]	91,9	87,8	89	88,8
Vada [%]	65,6	60,9	61,3	63
Průměr [%]	78,75	74,35	75,15	75,9
Příznaky	PLP, Δ PLP, ΔΔ PLP			
k	1	3	5	19
Bez vady [%]	85,9	84,6	87,6	85
Vada [%]	65,5	58,8	63,3	63,4
Průměr [%]	75,7	71,7	75,45	74,2
Příznaky	MFCC, PLP			
k	1	3	5	19
Bez vady [%]	83,2	89,4	89,2	86
Vada [%]	67	63,4	62,6	63,2
Průměr [%]	75,1	76,4	75,9	74,6
Příznaky	MFCC, Δ MFCC, PLP, Δ PLP			
k	1	3	5	19
Bez vady [%]	86,8	92,8	86	87
Vada [%]	66	69,4	65,2	61,8
Průměr [%]	76,4	81,1	75,6	74,4
Příznaky	MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP			
k	1	3	5	19
Bez vady [%]	88,6	86,2	86,2	89,4
Vada [%]	64,2	64,2	70,2	60
Průměr [%]	76,4	75,2	78,2	74,7

Tab. 8.4 Výsledky testů k -NN pro hlásku [š].

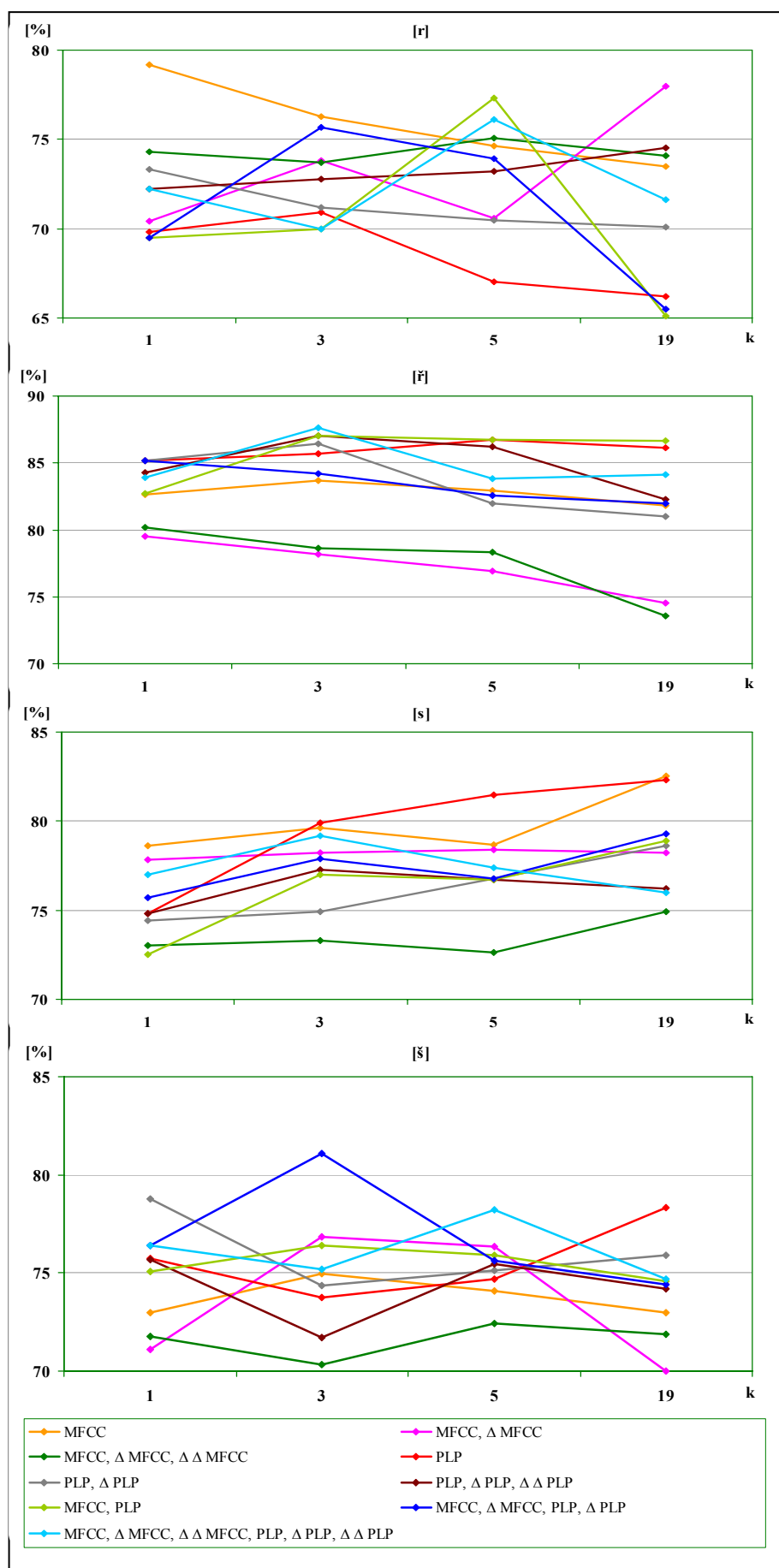
Data z tabulek 8.1 až 8.4 testování klasifikátoru k -NN jsou znázorněny v grafech v obrázku 8.2.

Z uvedených grafů vyplývá, že pro klasifikaci všech čtyř hlásek je nejvhodnější použití klasifikátor k -NN při $k = 3$. V následující tabulce jsou proto shrnuty výsledky pro toto nastavení.

[r]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	84,6	68,0	76,3
MFCC, Δ MFCC	77,2	70,4	73,8
MFCC, Δ MFCC, ΔΔ MFCC	73,2	74,2	73,7
PLP	74,2	67,6	70,9
PLP, Δ PLP	74,0	68,4	71,2
PLP, Δ PLP, ΔΔ PLP	80,0	65,6	72,8
MFCC, PLP	71,0	69,0	70,0
MFCC, Δ MFCC, PLP, Δ PLP	78,0	73,4	75,7
MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP	78,8	61,2	70,0

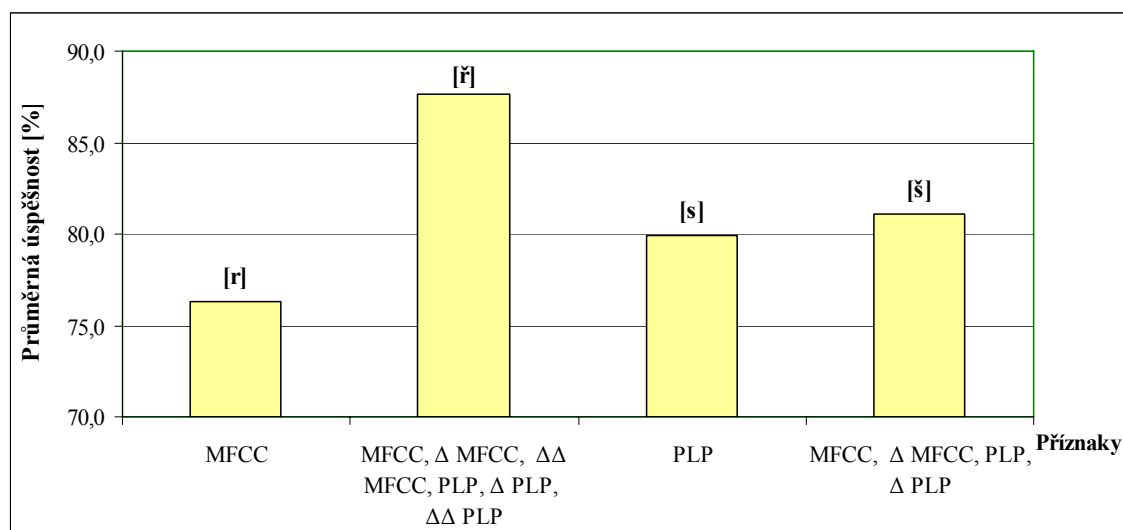
[ř]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	88,3	79,0	83,7
MFCC, Δ MFCC	83,8	72,5	78,2
MFCC, Δ MFCC, ΔΔ MFCC	90,3	67,0	78,7
PLP	88,9	82,5	85,7
PLP, Δ PLP	89,5	83,3	86,4
PLP, Δ PLP, ΔΔ PLP	90,2	83,8	87,0
MFCC, PLP	88,9	85,2	87,1
MFCC, Δ MFCC, PLP, Δ PLP	87,1	81,3	84,2
MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP	91,4	83,9	87,7
[s]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	82,5	76,7	79,6
MFCC, Δ MFCC	79,9	76,6	78,3
MFCC, Δ MFCC, ΔΔ MFCC	69,2	77,4	73,3
PLP	87,2	72,6	79,9
PLP, Δ PLP	84,8	65,0	74,9
PLP, Δ PLP, ΔΔ PLP	83,0	71,6	77,3
MFCC, PLP	83,6	70,4	77,0
MFCC, Δ MFCC, PLP, Δ PLP	89,6	66,2	77,9
MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP	87,2	71,2	79,2
[š]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	87,3	62,6	75,0
MFCC, Δ MFCC	88,6	65,1	76,9
MFCC, Δ MFCC, ΔΔ MFCC	90,3	50,4	70,4
PLP	84,2	63,3	73,8
PLP, Δ PLP	87,8	60,9	74,4
PLP, Δ PLP, ΔΔ PLP	84,6	58,8	71,7
MFCC, PLP	89,4	63,4	76,4
MFCC, Δ MFCC, PLP, Δ PLP	92,8	69,4	81,1
MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP	86,2	64,2	75,2

Tab. 8.5 Výsledky testu k -NN pro $k = 3$.



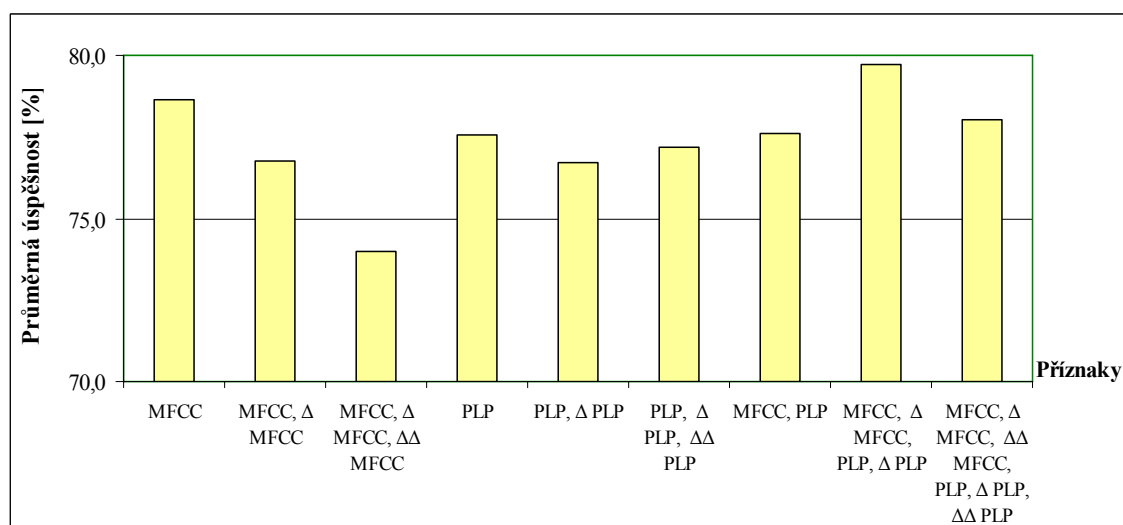
Obr. 8.2 Grafy vycházející z tabulek 8.1 až 8.4

Z tabulky 8.5 byly vybrány nejvyšší hodnoty průměrné úspěšnosti a tyto hodnoty jsou znázorněny v grafu 8.3.



Obr. 8.3 Graf nejvyšší úspěšnosti k -NN pro jednotlivé hlásky, $k = 3$.

Následující graf zobrazuje průměrnou úspěšnost klasifikace pro jednotlivé nastavení příznaků pro $k = 3$.



Obr. 8.4 Graf úspěšnosti k -NN v závislosti na kombinaci příznaků.

Z výsledků testování klasifikátoru k -NN vycházejí následující závěry:

Nejvyšší úspěšnosti správné klasifikace dosahuje klasifikátor k -NN u hlásky [ř] s nastavením kombinace příznaků MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP pro $k = 3$. Co se týče nastavení kombinace příznaků, pohybuje se úspěšnost klasifikace v průměru pro všechny hlásky mezi 74 % a 79,7 %. Nejvyšší úspěšnost klasifikace vychází pro kombinaci příznaků MFCC, Δ MFCC, PLP, Δ PLP pro $k = 3$.

8.5 Výsledky testování klasifikátoru SVM

Pro klasifikátor SVM při předběžných testech vycházely velice rozdílné a nízké hodnoty úspěšnosti klasifikace hlásek, proto již nebyl dále testován. Tato metoda není pro detekci logopedických vad v řeči vhodná.

8.6 Výsledky testování klasifikátoru ANN

Klasifikátor ANN byl testován s následujícím nastavením parametrů:

- počet vrstev a počet neuronů v jednotlivých vrstvách: 10-10-2.

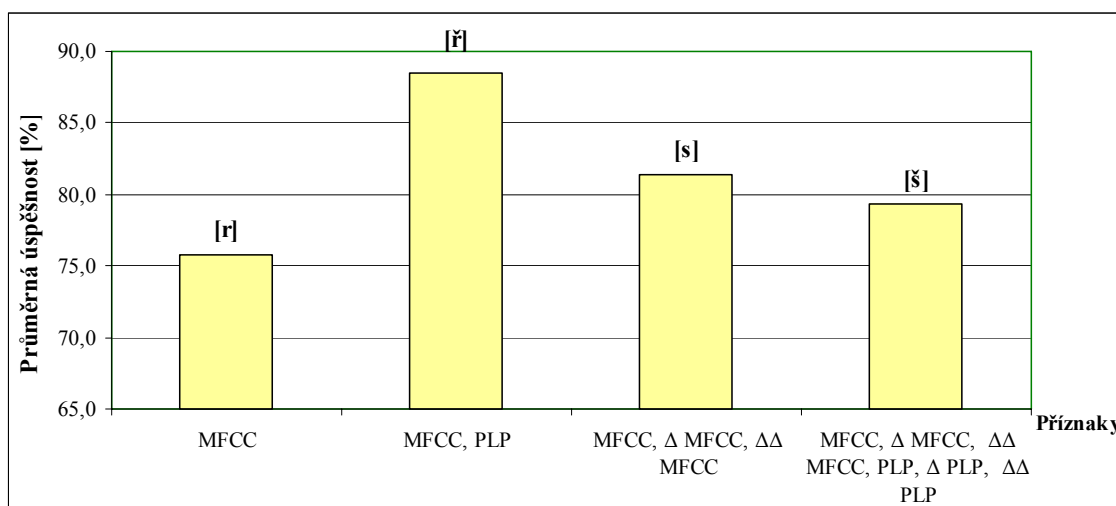
Tabulka 8.6 zobrazuje výsledky úspěšnosti klasifikace pomocí metody ANN pro jednotlivé hlásky a nastavení kombinace příznaků.

[r]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	78,8	72,8	75,8
MFCC, Δ MFCC	80,8	57,0	68,9
MFCC, Δ MFCC, $\Delta\Delta$ MFCC	83,0	58,5	70,8
PLP	85,0	63,0	74,0
PLP, Δ PLP	80,5	53,0	66,8
PLP, Δ PLP, $\Delta\Delta$ PLP	80,5	55,5	68,0
MFCC, PLP	78,0	67,0	72,5
MFCC, Δ MFCC, PLP, Δ PLP	75,0	58,5	66,8
MFCC, Δ MFCC, $\Delta\Delta$ MFCC, PLP, Δ PLP, $\Delta\Delta$ PLP	77,7	61,3	69,5
[ř]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	76,7	84,0	80,4
MFCC, Δ MFCC	85,7	81,3	83,5
MFCC, Δ MFCC, $\Delta\Delta$ MFCC	92,0	77,0	84,5
PLP	86,3	76,8	81,5
PLP, Δ PLP	69,0	75,0	72,0
PLP, Δ PLP, $\Delta\Delta$ PLP	88,0	81,0	84,5
MFCC, PLP	91,7	85,3	88,5
MFCC, Δ MFCC, PLP, Δ PLP	81,5	84,5	83,0
MFCC, Δ MFCC, $\Delta\Delta$ MFCC, PLP, Δ PLP, $\Delta\Delta$ PLP	81,5	83,0	82,3
[s]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	83,7	70,3	77,0
MFCC, Δ MFCC	85,7	68,3	77,0
MFCC, Δ MFCC, $\Delta\Delta$ MFCC	89,7	73,0	81,3
PLP	91,0	68,0	79,5
PLP, Δ PLP	87,0	71,0	79,0
PLP, Δ PLP, $\Delta\Delta$ PLP	85,7	76,0	80,8
MFCC, PLP	89,0	68,3	78,7
MFCC, Δ MFCC, PLP, Δ PLP	83,7	69,3	76,5
MFCC, Δ MFCC, $\Delta\Delta$ MFCC, PLP, Δ PLP, $\Delta\Delta$ PLP	69,0	85,0	77,0

[š]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	79,3	76,0	77,6
MFCC, Δ MFCC	72,7	78,0	75,3
MFCC, Δ MFCC, ΔΔ MFCC	80,5	67,8	74,1
PLP	76,8	79,4	78,1
PLP, Δ PLP	84,3	73,0	78,6
PLP, Δ PLP, ΔΔ PLP	80,0	74,0	77,0
MFCC, PLP	80,5	68,0	74,3
MFCC, Δ MFCC, PLP, Δ PLP	81,5	71,0	76,3
MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP	83,8	75,0	79,4

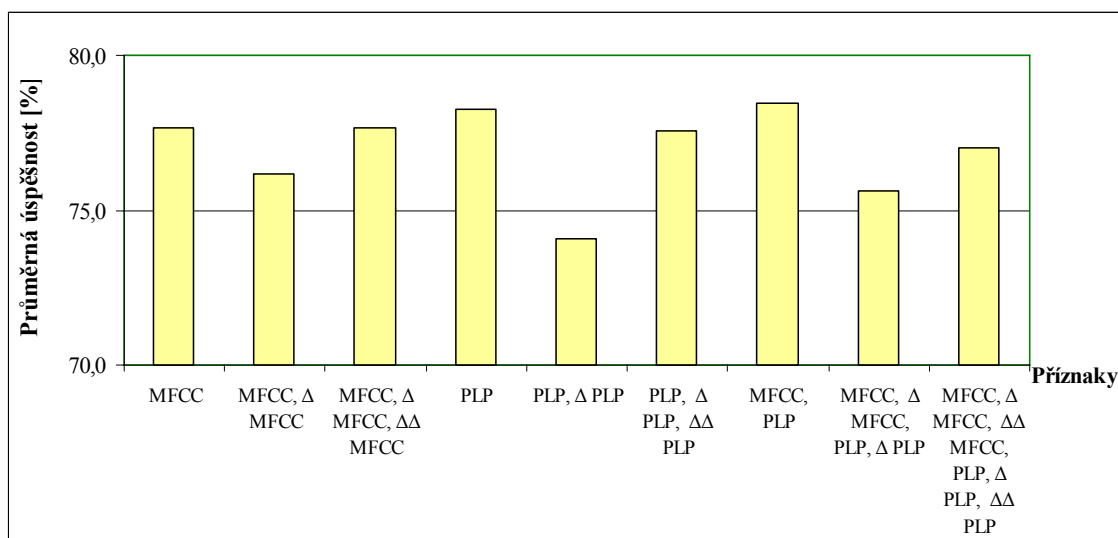
Tab. 8.6 Výsledky testu ANN

Následující obrázek zobrazuje nejvyšší hodnoty průměrné úspěšnosti klasifikace metodou ANN pro jednotlivé hlásky.



Obr. 8.5 Graf nevyšší úspěšnosti ANN pro jednotlivé hlásky.

Graf v obrázku 8.7 zobrazuje průměrnou úspěšnost klasifikace pro jednotlivé nastavení příznaků.



Obr. 8.6 Graf úspěšnosti ANN v závislosti na kombinaci příznaků.

Z výsledků testování klasifikátoru ANN vycházejí následující závěry:

Nejvyšší úspěšnosti správné klasifikace dosahuje klasifikátor ANN u hlásky [ř] s nastavením kombinace příznaků MFCC, PLP. V průměru pro všechny hlásky vychází úspěšnost klasifikace mezi 74,1 % a 78,5 %. Nejvyšší hodnoty je dosaženo pro kombinaci příznaků MFCC, PLP.

8.7 Výsledky testování klasifikátoru GMM

Pro testování GMM byly nastaveny následující parametry:

- počet Gaussových funkcí: 40.

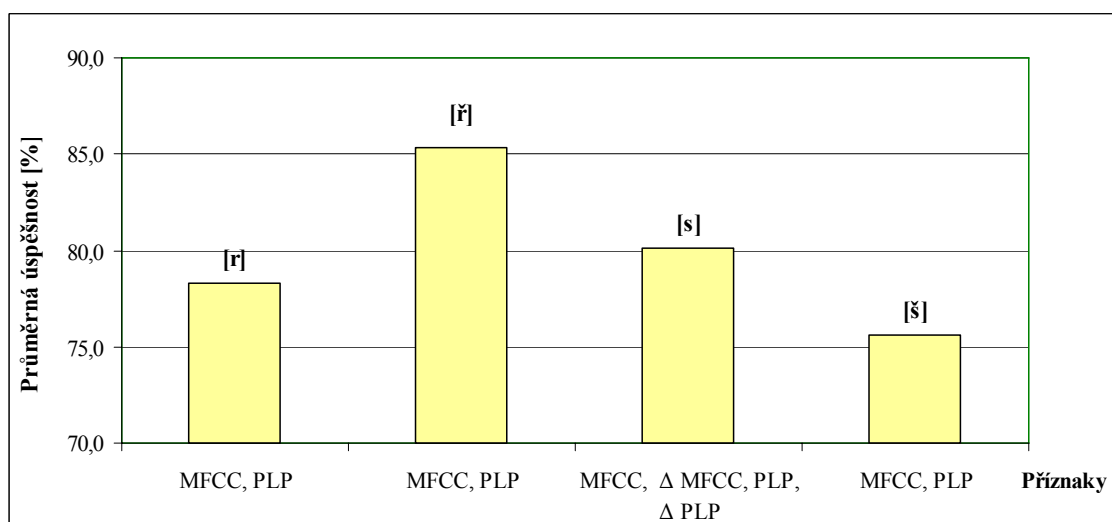
Výsledky testování klasifikátoru GMM obsahuje tabulka 8.7.

[r]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	83,1	60,7	71,9
MFCC, Δ MFCC	79,7	68,3	74,0
MFCC, Δ MFCC, ΔΔ MFCC	83,4	63,2	73,3
PLP	83,4	64,6	74,0
PLP, Δ PLP	84,5	68,9	76,7
PLP, Δ PLP, ΔΔ PLP	79,4	73,3	76,4
MFCC, PLP	86,5	70,1	78,3
MFCC, Δ MFCC, PLP, Δ PLP	82,4	72,3	77,4
MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP	80,2	69,9	75,1

[ř]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	75,2	86,2	80,7
MFCC, Δ MFCC	80,0	90,0	85,0
MFCC, Δ MFCC, ΔΔ MFCC	80,3	89,8	85,1
PLP	81,4	85,4	83,4
PLP, Δ PLP	78,0	86,0	82,0
PLP, Δ PLP, ΔΔ PLP	79,0	90,0	84,5
MFCC, PLP	85,3	85,4	85,4
MFCC, Δ MFCC, PLP, Δ PLP	84,3	82,4	83,4
MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP	79,0	88,0	83,5
[s]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	76,7	72,1	74,4
MFCC, Δ MFCC	75,1	70,2	72,7
MFCC, Δ MFCC, ΔΔ MFCC	73,8	70,7	72,3
PLP	85,0	65,0	75,0
PLP, Δ PLP	63,1	61,2	62,2
PLP, Δ PLP, ΔΔ PLP	64,2	61,2	62,7
MFCC, PLP	83,8	72,1	78,0
MFCC, Δ MFCC, PLP, Δ PLP	83,2	77,1	80,2
MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP	81,3	75,1	78,2
[š]			
Příznaky	Bez vady [%]	Vada [%]	Průměr [%]
MFCC	80,0	68,0	74,0
MFCC, Δ MFCC	75,8	65,1	70,5
MFCC, Δ MFCC, ΔΔ MFCC	75,3	70,1	72,7
PLP	80,0	63,0	71,5
PLP, Δ PLP	76,0	69,6	72,8
PLP, Δ PLP, ΔΔ PLP	76,5	70,5	73,5
MFCC, PLP	80,4	70,9	75,7
MFCC, Δ MFCC, PLP, Δ PLP	75,3	73,2	74,3
MFCC, Δ MFCC, ΔΔ MFCC, PLP, Δ PLP, ΔΔ PLP	77,4	70,5	74,0

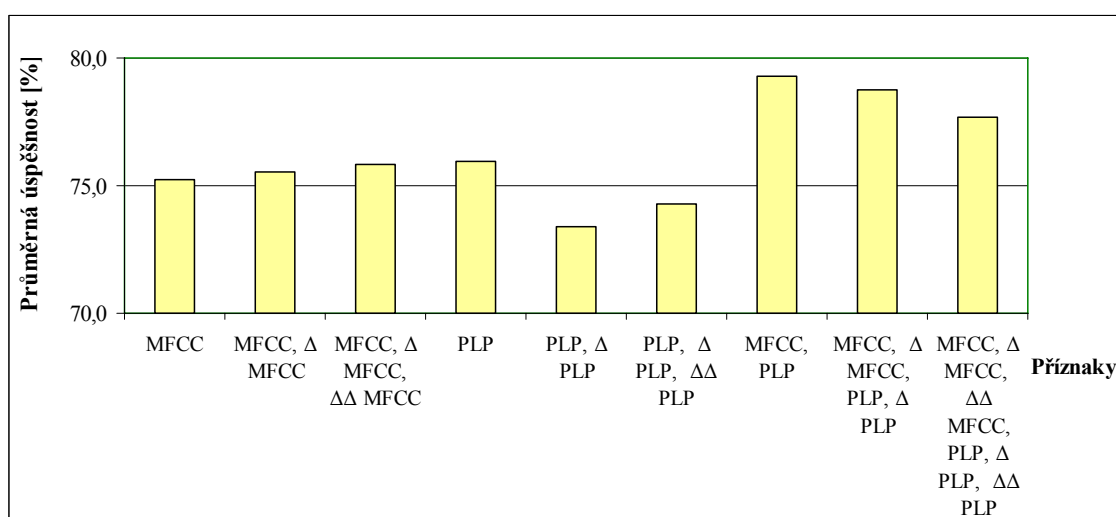
Tab. 8.7 Výsledky testu GMM.

Nejvyšší hodnoty průměrné úspěšnosti klasifikace metodou GMM pro jednotlivé hlásky se zvolenou kombinací příznaků zobrazuje Obr. 8.7.



Obr. 8.7 Graf nevyšší úspěšnosti GMM pro jednotlivé hlásky

Průměrná úspěšnost klasifikace pro jednotlivé nastavení příznaků je zobrazena v Obr. 8.8.



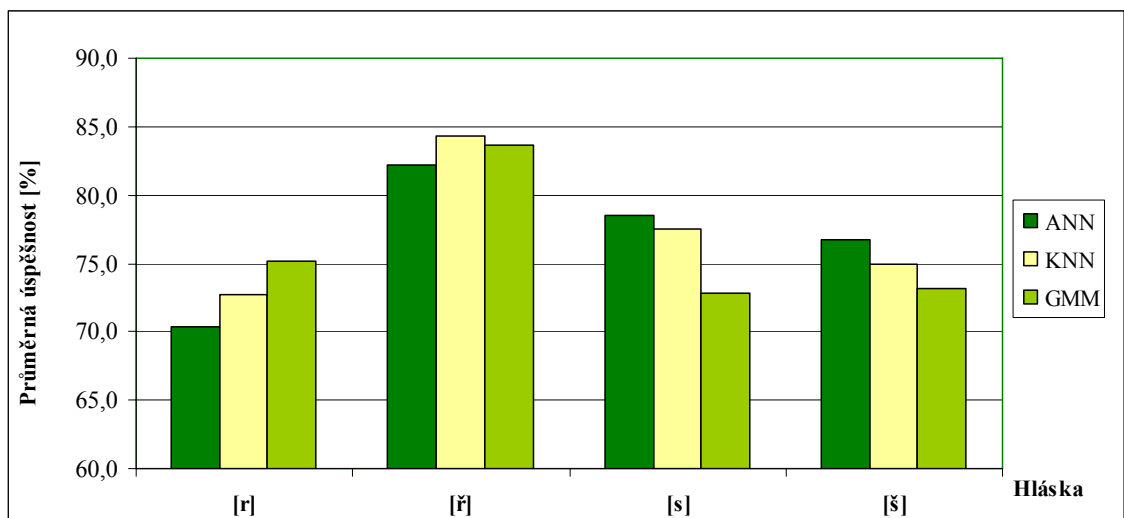
Obr. 8.8 Graf úspěšnosti GMM v závislosti na kombinaci příznaků.

Z výsledků testování klasifikátoru GMM vycházejí následující závěry:

Nejvyšší úspěšnosti správné klasifikace dosahuje klasifikátor GMM u hlásky [ř] s nastavením kombinace příznaků MFCC, PLP. Při změně kombinace příznaků vychází úspěšnost klasifikace v průměru pro všechny hlásky mezi 73,4 % a 79,3 %. Nejvyšší hodnoty je dosaženo pro kombinaci příznaků MFCC, PLP.

8.8 Celkové výsledky testovaných klasifikátorů

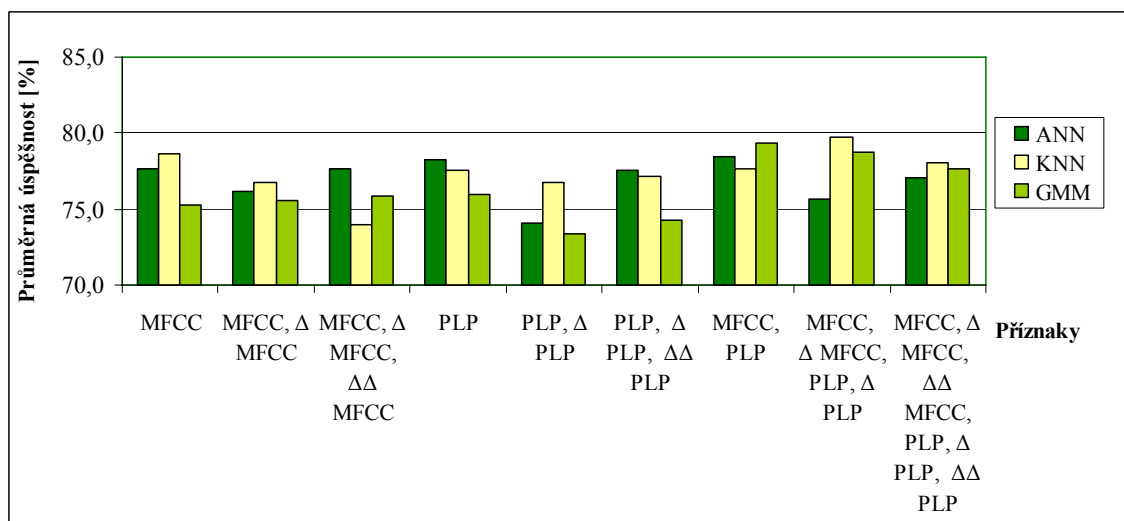
Graf v Obr. 8.9 znázorňuje celkovou průměrnou úspěšnost klasifikátorů k -NN, ANN a GMM pro jednotlivé hlásky.



Obr. 8.9 Celková úspěšnost klasifikátorů k -NN, ANN a GMM.

Všechny tři metody dosahují podobných výsledků úspěšnosti klasifikace u všech hlásek, a to pro [r] v rozmezí 70,3 – 75,2 %, pro [ř] v rozmezí 82,2 – 84,3 %, pro [s] v rozmezí 72,8 – 78,5 %, a pro [š] v rozmezí 73,2 – 76,7.

Obr. 8.10 zobrazuje celkovou úspěšnost klasifikace všech tří testovaných klasifikátorů pro jednotlivé kombinace příznaků.



Obr. 8.10 Úspěšnost klasifikátorů pro jednotlivé kombinace příznaků.

Všechny testované kombinace dosahují podobných hodnot v rozmezí od 73,4 - 79,7 %. Nejlepších průměrných hodnot dosahuje klasifikace kombinace příznaků MFCC, PLP, a to 78,5 %.

Výsledky testování vycházejí pro všechny metody klasifikace s různými kombinacemi příznaků s malým rozdílem cca 5 %, s výjimkou klasifikátoru SVM jehož, použití bylo pro detekci logopedických vad v řeči zamítnuto již na začátku testování.

Pro konečnou aplikaci jsou vybrány následující metody:

Pro segmentaci slova na hlásky je použita metoda rozdílnosti příznaků, kde jsou příznaky výstupem mel-banky a hláska je vybrána pomocí metody GMM (viz. kapitola 6.2.) Příznaky pro následnou klasifikaci jsou extrahovány metodou MFCC s 20 příznaky, délka rámce 256 (16 ms) s přesahem 128 (8 ms). Výběr příznaků je proveden metodou klasifikace rámeček po rámci a vybrané příznaky jsou klasifikovány metodou ANN.

Závěr

Cílem této práce bylo navrhnout a realizovat metodu pro rozpoznání logopedických vad v řeči a prozkoumat metody výpočtu příznaků a klasifikace pro tento účel.

Pro výpočet příznaků z řeči, které jsou následně použity pro klasifikaci, byly prozkoumány metody MFCC a PLP, pro klasifikaci příznaků metody k -NN, SVM, ANN a GMM. Tyto metody byly v textu popsány a následně testovány.

Dále byl navržen algoritmus pro segmentaci vysloveného slova a pro výběr hlásky, která má být klasifikována. Pro detekci logopedických vad v řeči byl v programovém prostředí Matlab navržen a realizován algoritmus, který jednou z metod extrakce příznaků vypočítá příznaky a tyto příznaky jednou z metod klasifikace klasifikuje do tříd vadné nebo správné hlásky.

Navržené algoritmy byly závěrem práce testovány pro hlásky [r], [ř], [s] a [š]. Výsledky testů jsou zaznamenány v tabulkách a zobrazeny v jednotlivých grafech. Úspěšnost algoritmu pro segmentaci slova se pohybuje kolem 70 %. Výběr hlásky je úspěšný pouze pro hlásky [s] a [š] (cca 90 %), pro hlásky [r] a [ř] se úspěšnost pohybovala pod 50 %. Pro zlepšení výsledků úspěšnosti by bylo možné vytvořit kvalitnější databázi trénovacích nahrávek, případně použít jinou metodu. Následně byla testována úspěšnost klasifikace příznaků, kdy byly příznaky vypočteny různou kombinací metod PLP, MFCC a jejich první a druhou derivací. Jako klasifikátory byly testovány pouze metody k -NN, ANN a GMM. Klasifikátor SVM při předběžných testech nedosahoval uspokojivých hodnot, proto nebyl dále testován. Výsledky úspěšnosti klasifikace metodami k -NN, ANN a GMM dosahovaly přibližně stejných hodnot +/- 5 %. Nejlépe byla klasifikována hláska [ř] (cca 83 %), nejhorší úspěšnosti dosáhla klasifikace hlásky [r] (cca 73 %). Změna kombinací metod extrakce příznaků neměla na úspěšnost klasifikace podstatný vliv, průměrná úspěšnost se měnila v rozmezí 6 %.

Jako konečná podoba navrženého algoritmu byla vytvořena uživatelská aplikace, která byla graficky přizpůsobena dětem. Pro použití aplikace v praxi by musela být zdokonalena metoda pro výběr hlásky ze segmentovaného slova a vytvořena kvalitní databáze trénovacích nahrávek.

Použitá literatura

- [1] NEJEZCHLEBOVÁ, Lenka. Děti stále více šišljají, můžou za to i počítače. *IDNES.cz* [online]. 2008 [cit. 2008-12-16]. Dostupný z WWW: <http://zpravy.idnes.cz/deti-stale-vice-sislaji-muzou-za-to-i-pocitace-fo1-domaci.asp?c=A080406_224417_domaci_dp>.
- [2] PSUTKA, Josef, et al. *Mluvíme s počítačem česky*. Aleš Baďura. 1. vyd. Praha : Academia, 2006. 752 s. ISBN 80-200-1309-1.
- [3] KLENKOVÁ, Jiřina. *Logopedie*. Mgr. Alena Herbergová. 1. vyd. Praha 7 : GRADA, 2006. 228 s. ISBN 80-247-1110-9.
- [4] NEUBAUER, Karel. *Logopedie : učební text pro bakalářské studium speciální pedagogiky*. 2. přeprac. rozšíř. vyd. Hradec Králové : GAUDEAMUS, 2007. 107 s. ISBN 978-80-7041-093-6.
- [5] VITÁSKOVÁ, Kateřina, PEUTELSCHMIEDOVÁ, Alžběta. *Logopedie*. RNDr. Hana Dziková. 1. vyd. Olomouc : Univerzita Palackého, 2005. 182 s. ISBN 80-244-11088-5.
- [6] HERMANŠKY, H., *Perceptual Linear Predictive (PLP) Analysis of Speech*. J. Acoust. Soc. Am., 87, č.4, 1990, s. 1738 – 1752.
- [7] ELLIS, Daniel P. W.. *PLP and RASTA (and MFCC, and inversion) in Matlab* [online]. 2005- , 2005 [cit. 2008-10-29]. Text v angličtině. Dostupný z WWW: <<http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/>>.
- [8] MOLNÁR, Karol. Úvod do problematiky umělých neuronových sítí. *Elektrorevue* [online]. 2000 [cit. 2009-05-09]. Dostupný z WWW: <<http://www.elektrorevue.cz/clanky/00013/index.html>>.
- [9] JIRSÍK, Václav, HRÁČEK, Petr. *Neuronové systémy, expertní systémy a rozpoznávání řeči*. Brno : [s.n.], 2000. 106 s. Skriptum FEKT VUT v Brně.
- [10] BUČEK, Jindřich. Segmentace řeči na fonémy metodou sledování rozdílů příznaků. In *Sborník prací studentů a doktorandů*. Brno : Akademické nakladatelství CERM, 1997. Ústav radioelektroniky. s. 73-75. ISBN 80-214-0637-2.
- [11] ESPOSITO, Anna, AVERSANO, Guido. *Text Independent Methods for Speech Segmentation.*, s. 34.
- [12] DUŠIL, Lubomír. *Automatické rozpoznávání logopedických vad v řečovém projevu*. [s.l.], 2008. 48 s. Vedoucí semestrální práce prof. Ing. Zdeněk Směkal, CSc.
- [13] ZAPLATÍLEK, Karel, DOŇAR, Bohuslav. *Matlab : tvorba uživatelských aplikací*. Iveta Kubicová. 1. vyd. Praha : BEN, 2005. 216 s. ISBN 80-7300-133-0.

Seznam zkratek

- ANN – Umělé neuronové sítě (Artificial neural network)
- DCT – Diskrétní kosinová transformace (Discrete Cosine Transform)
- FFT – Rychlá Fourierova transformace (Fast Fourier Transform)
- GMM – Smíšené Gaussovy modely (Gaussian Mixture Models)
- IDFT – Zpětná diskrétní Fourierova transformace (Inverse Discrete Fourier Transform)
- KKP – Kognitivně Komunikační Poruchy
- k -NN – Algoritmus k -nejbližších sousedů (k -Nearest Neighbors algorithm)
- LDA – Lineární diskriminační analýza (Linear Diskriminant Analysis)
- MFCC – Melovské keprstrální koeficienty (Mel-Frequency Cepstral Coefficients)
- PCA – Analýza hlavních komponent (Principal Component Analysis)
- PLP – Perceptivní lineární predikce (Perceptual Linear Prediction)
- RBF – Metoda Radiálních Bázových Funkcí (Radial Basis Function)
- SVM – Algoritmy podpůrného učení (Support Vector Machines)
- Δ MFCC – první derivace MFCC
- Δ PLP – první derivace PLP
- $\Delta\Delta$ MFCC – druhá derivace MFCC
- $\Delta\Delta$ PLP – druhá derivace PLP

Seznam symbolů

- $\Psi(z)$ – frekvenční charakteristika pásmové propusti pro PLP
- μ_i^s – střední hodnota pravděpodobnosti
- Φ_m – přizpůsobení kritických pásmových filtrů křivkou stejné hlasitosti
- A – antiformant
- a_i - koeficienty číslicového filtru LPC
- $B(j)$ – křivka rozdílnosti příznaků
- $c[n]$ – keprstrum signálu
- C_i^s – kovarianční matice
- D – procentuální vyjádření nadsegmentování/podsegmentování
- E – odchylka nalezených hranic hlásky od skutečných

$E(\omega)$ – křivky stejné hlasitosti

F – formant

f – frekvence [Hz]

F_0 - frekvence základního tónu

f_m – frekvence [mel]

G - koeficient zesílení

$H(z)$ - přenosová funkce

h_d - poloha detekované hranice hlásky

h_s - poloha skutečné hranice hlásky

k – počet nejbližších sousedů algoritmu k -NN

M^s – počet Gaussových funkcí rozložení pravděpodobnosti

$p(o|\lambda^s)$ – vážená lineární kombinace hustotních funkcí normálních rozdělení jednotlivých tříd

$P(\omega)$ - krátkodobé výkonové spektrum signálu

P_c – procentuální počet správně detekovaných hranic

P_m - procentuální počet chybně detekovaných hranic

Q - řád modelu LPC

$S(\omega)$ – spektrum signálu

$s[n]$ – vstupní signál

S_c – počet správně detekovaných hranic ve slově

S_{ins} – nadsegmentování/podsegmentování

S_t – počet skutečných hranic ve slově

T_0 - perioda základního hlasivkového tónu

$u(f,i)$ – trojúhelníkový filtr mel-banky

$u[n]$ – buzení signálu

Θ - vážená spektrální sumarizace

$\Omega(\omega)$ - osa frekvencí [bark]

ω - osa frekvencí [rad/s]

Obsah přiloženého CD

- Elektronická verze diplomové práce,
- program Učíme se správně mluvit,
- nástroj na testování úspěšnosti segmentace slova na hlásky.